8. Networking and Content Delivery

Amazon API Gateway

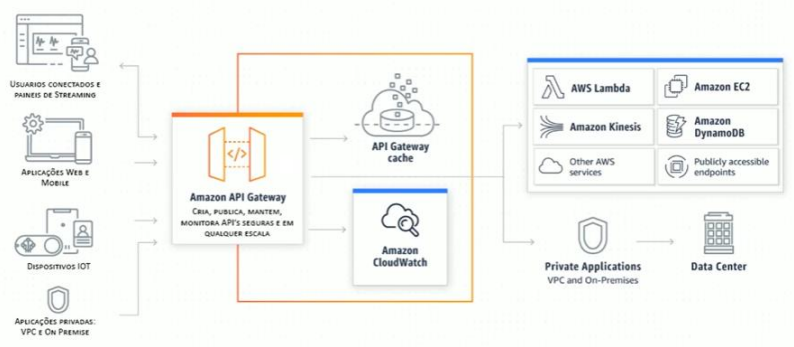
O AWS API Gateway é um serviço gerenciado que permite que desenvolvedores criem, publiquem, mantenham, monitorem e protejam APIs em qualquer escala com facilidade

Usando o API Gateway, é possível criar APIs RestFul e APIs WebSockets que habilitam aplicativos de comunicação bidirecionais em tempo real

O API Gateway dá suporte a cargas de trabalho conteinerizadas (ex: EC2) e sem servidor, além de aplicativos da web

Ele administra todas as tarefas envolvidas no recebimento e processamento de até centenas de milhares de chamadas de API simultâneas

Inclusive gerenciamento de tráfego, suporte de CORS, controle de autorização e acesso, com fluxo controlado, monitoramento e gerenciamento de versões de API



Além disso, com o modelo de definição de preço em camadas do API Gateway, é possível reduzir custos à medida que seu uso da API é escalado (quanto mais é usado, mais barato fica)

Existem 2 tipos de APIs:

1. **APIs RESTful:** Existem dois tipos de APIs RESTful, o primeiro é o proxy HTTP que é um serviço de proxy simples para repassar as chamadas aos nossos serviços. O segundo sendo a API REST, que garante funcionalidade de proxy de API e recursos de gerenciamento de API em uma única solução
2. **APIs WebSockets:** Crie aplicativos de comunicação bidirecionais em tempo real, como aplicativos de bate-papo e painéis de streaming, com APIs WebSocket

Cache do API Gateway: O API Gateway também oferece a opção de armazenar dados em cache mediante uma taxa por hora que varia de acordo com o tamanho de cache selecionado. Ou seja, ao invés da API Gateway executar um lambda que busca o mesmo dado no S3, estes dados ficam cacheados, reduzindo chamadas a serviços de forma desnecessária

O API Gateway não tem taxas mínimas ou custos antecipados. Você paga apenas pelas chamadas de API recebidas e pela quantidade transferida de dados de saída

Para APIs HTTP e APIs Rest, o pagamento ocorre apenas pelas chamadas de API recebidas e pela quantidade de dados transferida para fora

Para APIs do WebSocket, o pagamento ocorre somente quando as APIs estão em uso com base no número de mensagens enviadas e recebidas e pelos minutos de conexão

O nível gratuito do API Gateway inclui 1 milhão de chamadas de API HTTP, 1 milhão de chamadas de API Rest, 1 milhão de mensagens e 750.000 minutos de conexão por mês durante até 12

Com um preço de solicitações de API de apenas 0,90 USD a cada milhão de solicitações na camada mais alta, você pode reduzir seus custos conforme seu uso de API aumenta por região nas suas contas da AWS.

Link de definição de preços API Gateway: https://aws.amazon.com/pt/api-gateway/pricing/

Não esquecer dos custos adicionais de outros serviços que são usados em conjunto como o AWS Lambda e AWS CloudWatch

Controle de Utilização

O Amazon API Gateway oferece controle de utilização em diversos níveis, incluindo os níveis global e por serviço. Os limites do controle de utilização podem ser definidos para taxas padrão e picos. Por exemplo, os proprietários de uma API podem definir um limite de taxa de 1.000 solicitações por segundo para um método específico nas APIs REST, bem como configurar o Amazon API Gateway para processar picos de 2.000 solicitações por segundo durante alguns segundos. O Amazon API Gateway controla o número de solicitações por segundo. Todas as solicitações acima do limite receberão uma resposta **HTTP 429**. Os SDKs cliente (exceto o Javascript) gerados pelo Amazon API Gateway repetem automaticamente as chamadas quando recebem essa resposta.

O controle de utilização garante o controle do tráfego de APIs para ajudar os serviços de back-end a manter o desempenho e a disponibilidade.

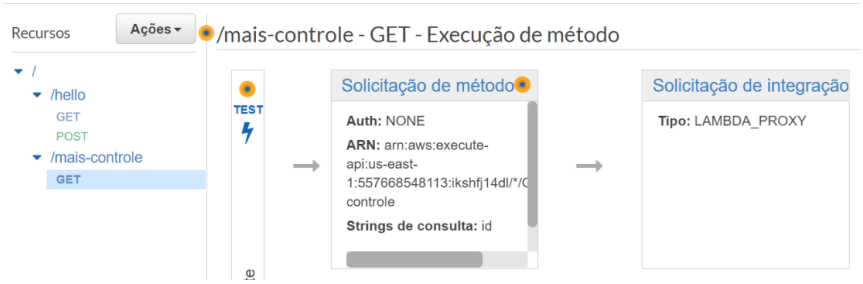
Os limites de taxa de controle de utilização podem ser definidos por método. Você pode editar os limites de controle de utilização nas configurações do método usando as APIs do Amazon API Gateway ou o console do Amazon API Gateway.

Compreendendo API Gateway Lambda Proxy Integration

https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html

https://docs.aws.amazon.com/apigateway/latest/developerguide/set-up-lambda-proxy-integrations.html#api-gateway-simple-proxy-for-lambda-input-format

Para acessar a opção Lambda Proxy Integration, clicar em **Solicitação de Integração**

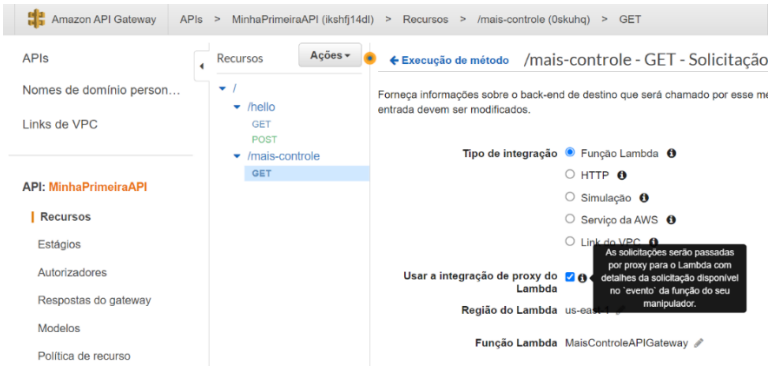


A ideia é que a integração com o proxy lambda transforme o formato da solicitação, ou seja, em vez de a resposta vir no formato de String como um único body, os campos do body são retornados como uma requisição. Exemplo:

```
{
  "isBase64Encoded": true|false,
  "statusCode": httpStatusCode,
  "headers": {“headerName": "headerValue", ...},
  "multiValueHeaders": {“headerName": ["headerValue", "headerValue2", ...], ...},
  "body": "..."
}
```

Dessa maneira, é transferido ao Lambda a responsabilidade de formatar a resposta da solicitação, dando mais controle sobre os endpoints criados quando integrados ao lambda

Na opção “Solicitação de Integração”, selecionar a opção de integração de proxy do lambda para capturar as informações da solicitação disponível

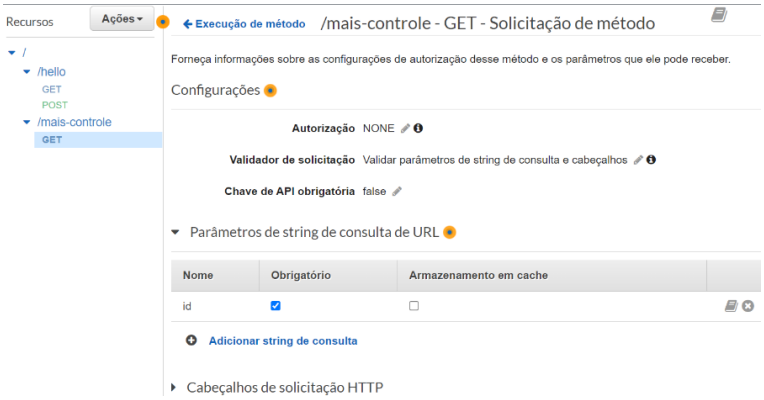


Autorizadores

Os autorizadores permitem que você controle o acesso às APIs usando os grupos de usuários do Amazon Cognito ou uma função Lambda.

Solicitação de Método

Na solicitação de método, é possível validar parâmetros de String, cabeçalhos de solicitação Http e o corpo da solicitação, por exemplo criar um parâmetro de String obrigatório, caso não seja passado na requisição, a própria API Gateway, e não o Lambda, irá retornar mensagem e status erro.



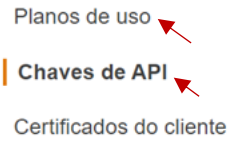
Chave de API

A opção **“Chave de API obrigatória”** tem uma grande importância, e serve como por exemplo para obrigar o usuário solicitante informar uma chave de API e a partir disso limitar a quantidade de requisições/segundo e a quantidade total mensal, além é claro, de garantir uma segurança adicional na chamada da API

Após mudar para true a opção “Chave de API obrigatória”, clicar na opção “Chaves de API”, após criar a chave, adicionar a um plano de uso, porém deve ir à opção “Planos de Uso”

Os planos de uso ajudam a medir o uso de APIs. Com eles, é possível aplicar um limite de cota e de controle de utilização em cada chave de API. Os limites de controle de utilização definem o número máximo de solicitações por segundo disponíveis para cada chave. Os limites de cota definem o número de solicitações que cada chave de API pode fazer em um período

Observação: não esquecer de implantar a API novamente para que as funcionalidades criadas tenham efeito e no header da request passar a chave como **x-api-key** e valor sendo a Chave de API



Control Access

You control access to Amazon API Gateway with IAM permissions by controlling access to the following two API Gateway component processes.

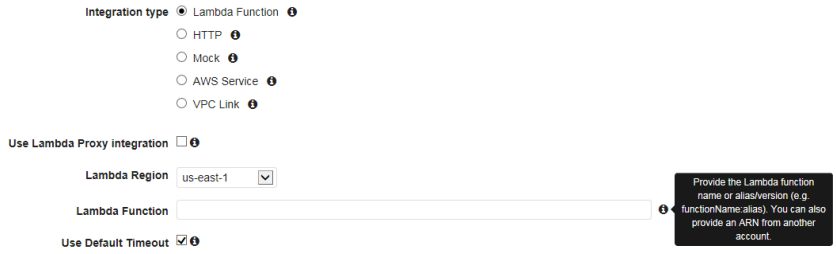
- To create, deploy, and manage an API in API Gateway, you must grant the API developer permissions to perform the required actions supported by the API management component of API Gateway.

- To call a deployed API or to refresh the API caching, you must grant the API caller permissions to perform required IAM actions supported by the API execution component of API Gateway.

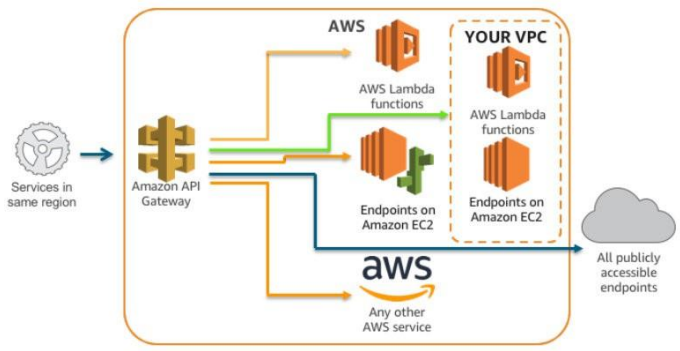
API Gateway Integration Type

Q: With what backends can Amazon API Gateway communicate?

Amazon API Gateway can execute AWS Lambda functions in your account, start AWS Step Functions state machines, or call HTTP endpoints hosted on AWS Elastic Beanstalk, Amazon EC2, and also **non-AWS hosted HTTP based operations that are accessible via the public Internet**. API Gateway also allows you to specify a mapping template to generate static content to be returned, helping you mock your APIs before the backend is ready. You can also integrate API Gateway with other AWS services directly. For example, you could expose an API method in API Gateway that sends data directly to Amazon Kinesis.



AWS API gateway can connect to AWS services, making proxy calls only to their respective AWS APIs. There is no integration type for database connections directly from API Gateway. You can use the Lambda function to connect with the database and make Lambda as an integration type for API Gateway. AWS has introduced VPC Link, a way to connect to the resources within a private VPC



Throttling limits

To help understand these throttling limits, here are a few examples, given the burst limit and the default account-level rate limit:

- If a caller submits 10,000 requests in a one-second period evenly (for example, 10 requests every millisecond), API Gateway processes all requests without dropping any.
- If the caller sends 10,000 requests in the first millisecond, API Gateway serves 5,000 of those requests and throttles the rest in the one-second period.

Controlling access to your AWS API Gateway

Controlling Access to an API in API Gateway

API Gateway supports multiple mechanisms for controlling access to your API:

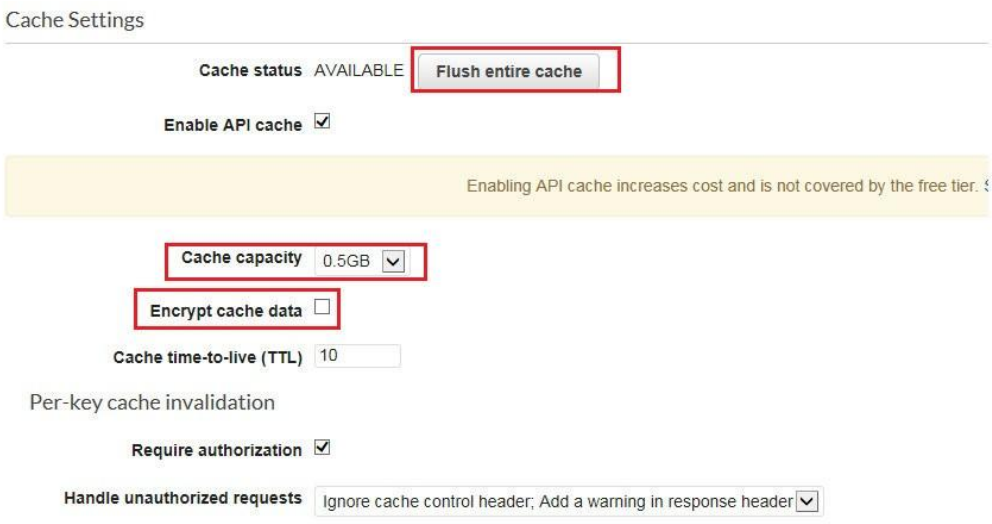
- **Resource policies** let you create resource-based policies to allow or deny access to your APIs and methods from specified source IP addresses or VPC endpoints.
- **Standard AWS IAM roles and policies** offer flexible and robust access controls that can be applied to an entire API or individual methods.
- **Cross-origin resource sharing (CORS)** lets you control how your API responds to cross-domain resource requests.
- **Lambda authorizers** are Lambda functions that control access to your API methods using bearer token authentication as well as information described by headers, paths, query strings, stage variables, or context variables request parameters.
- **Amazon Cognito user pools** let you create customizable authentication and authorization solutions.
- **Client-side SSL certificates** can be used to verify that HTTP requests to your backend system are from API Gateway.
- **Usage plans** let you provide API keys to your customers — and then track and limit usage of your API stages and methods for each API key.

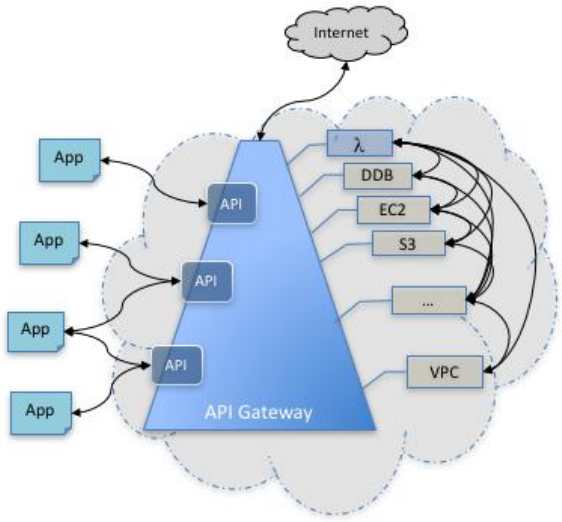
Security Measures By Default

API Gateway supports throttling settings for each method or route in your APIs. You can set a standard rate limit and a burst rate limit per second for each method in your REST APIs and each route in WebSocket APIs. **Further, API Gateway automatically protects your backend systems from distributed denial-of-service (DDoS) attacks, whether attacked with counterfeit requests (Layer 7) or SYN floods (Layer 3).**

Cache Settings

Following are the settings when enabling/disabling API caching for API Gateway.





➢ Pontos de Atenção

1. Você está usando uma combinação de API Gateway e Lambda para os serviços da web de seu portal online que está sendo acessado por centenas de milhares de clientes todos os dias. Sua empresa estará anunciando um novo produto revolucionário e espera-se que seu portal receba um grande número de visitantes em todo o mundo. Como você pode proteger seus sistemas de back-end e aplicativos de picos de tráfego? **R: Use throttling limits in API Gateway**

2. While managing permissions for the API Gateway, what could be used to ensure that the right level of permissions is given to Developers, IT Admins, and users? Also, the permissions should be easily managed. R: Use IAM Policies to create different policies for different types of users

3. Which of the following are valid integration sources for API Gateway? R: 1) Public facing HTTP-based endpoints outside AWS network. 2) Lambda functions from another account. 3) VPC Link

4. A Company ABC has 100 REST APIs exposed to the Internet from their on-premise network. They have already integrated with AWS through DirectConnect. They have approached you asking for a cost-effective way of making these REST APIs available through AWS API Gateway because of the resiliency and cost reductions provided by it. What solution would you provide? R: Use VPC Link to integrate on-premises backend solutions through DirectConnect and private VPC

5. You have built a REST API using API gateway and distributed to your customers. However, your API is receiving la rge number of requests and overloading your backend system causing performance bottlenecks and eventually c ausing delays and failures in serving the requests for your important customers. How would you improve the API performance? (Choose 2 options) R: 1) Enable throttling and control the number of requests per second 2) Enable API caching to serve frequently requested data from API cache.

6. You have created a public-facing REST API using AWS API Gateway with a default throttle setting of 10000 requests per second and a burst of 5000 requests. You are getting 8000 requests in the first millisecond. Which of the following statements is true? R: 5000 requests would succeed and throttles the rest of 3000 in the one-second period

7. In AWS API Gateway, which of the following security measures is provided default by AWS to protect the backend systems? R: Protection from distributed denial-of-service (DDoS) attacks

8. You have created a REST API using AWS API Gateway and deployed it to production. Your organization requested the details regarding who is accessing the API deployed to production. How would you get the required information? R: Enable Access Logging in CloudWatch API logging. because**,** in access logging, you, as an API

developer, want to log who has accessed your API and how the caller accessed the API. You can create your own log group or choose an existing log group that could be managed by API Gateway.

Amazon CloudFront

CloudFront é o serviço responsável por disponibilizar CDN (Content Delivery Network)

O Amazon CloudFront é um serviço rápido de rede de entrega de conteúdo (CDN) que entrega dados, vídeos, aplicações, APIs e Web Sites a clientes em todo o mundo com segurança, baixa latência e altas velocidades de transferência em um ambiente de uso facilitado para desenvolvedores.

Suponha que existe um arquivo de vídeo em uma AZ no Brasil, porém este conteúdo está sendo acessado por usuários de outras Regiões, para que o conteúdo não tenha delay e haja qualidade na entrega do stream de vídeo, o conteúdo é copiado então para a AZ (e também para as Edge Location) mais próxima da requisição realizada (apenas uma requisição já é o suficiente para a réplica do conteúdo)

Quando a cópia é realizada, um cache é criado para o conteúdo com uma validade (Lifetime), ou seja, o conteúdo fica cacheado por exemplo por 24h, após esse período o conteúdo é removido, caso ocorra uma nova requisição o processo se repete, ou seja, copia do conteúdo, cache do conteúdo e expiração do mesmo

Obviamente que a primeira requisição é lenta, pois haverá a copia do conteúdo, porém as próximas requisições serão bem mais rápidas pois estarão em cache na AZ mais próxima do requisitante

Os casos de uso mais comuns são de empresas que precisam agilizar o acesso a Web Sites, realizar Broadcast (transmissão) de vídeos e Live Streaming com delay reduzido

O que é o Amazon CloudFront?

O Amazon CloudFront é um serviço da web que acelera a distribuição do conteúdo estático e dinâmico da web, como arquivos .html, .css, .js e arquivos de imagem, para os usuários.

O CloudFront distribui seu conteúdo por meio de uma rede global de datacenters denominados pontos de presença. Ao solicitar um conteúdo que você está veiculando com o CloudFront, o usuário é roteado para o ponto de presença com a menor latência (atraso de tempo) para que o conteúdo seja fornecido com o melhor desempenho possível.

> Se o conteúdo já estiver no ponto de presença com a menor latência, o CloudFront o entregará imediatamente.
>
> Se o conteúdo não estiver nesse ponto de presença, o CloudFront recuperará de uma origem que você definiu—como um bucket do Amazon S3, um canal do MediaPackage ou um servidor HTTP (por exemplo, um servidor web), que você identificou como a fonte para a versão definitiva do seu conteúdo

Por exemplo, suponha que você esteja exibindo uma imagem de um servidor web tradicional, e não do CloudFront. Por exemplo, você pode fornecer uma imagem, sunsetphoto.png, usando a URL http://example.com/sunsetphoto.png.

Seus usuários podem navegar facilmente para esse URL e ver a imagem. Mas provavelmente não sabem que sua solicitação foi roteada de uma rede para outra—por meio do conjunto complexo de redes interconectadas que compõem a Internet—até a imagem ser encontrada.

O CloudFront acelera a distribuição do seu conteúdo encaminhando cada pedido de usuário por meio da rede backbone da AWS para o ponto de presença que veicule melhor seu conteúdo. Normalmente, esse é um servidor de ponto do CloudFront que fornece a entrega mais rápida ao visualizador.

Usar a rede da AWS reduz drasticamente o número de redes pelas quais as solicitações dos usuários devem passar, melhorando o desempenho. Os usuários obtêm menos latência (o tempo que leva para carregar o primeiro byte do arquivo) e taxas de transferência de dados maiores.

Você também pode obter mais confiabilidade e disponibilidade porque as cópias de seus arquivos (também conhecidos como *objetos*) agora são mantidos (ou armazenados em cache) em vários pontos de presença em todo o mundo.

Definição de Preço

Nível Gratuito: Após o cadastramento, os novos clientes da AWS recebem 50 GB de transferência de dados para fora e 2 milhões de solicitações HTTP e HTTPS, além de 2 milhões de invocações do CloudFront Functions por mês durante um ano.

Por mês	Estados Unidos, México e Canadá	Europa e Israel	África do Sul, Quênia e Oriente Médio	América do Sul	Japão	Austrália e Nova Zelândia	Hong Kong, Indonésia, Filipinas, Singapura, Coreia do Sul, Taiwan e Tailândia	Índia
Primeiros 10 TB	0,085 USD	0,085 USD	0,110 USD	0,110 USD	0,114 USD	0,114 USD	0,120 USD	0,109 USD
Próximos 40 TB	0,080 USD	0,080 USD	0,105 USD	0,105 USD	0,089 USD	0,098 USD	0,100 USD	0,085 USD
Próximos 100 TB	0,060 USD	0,060 USD	0,090 USD	0,090 USD	0,086 USD	0,094 USD	0,095 USD	0,082 USD
Próximos 350 TB	0,040 USD	0,040 USD	0,080 USD	0,080 USD	0,084 USD	0,092 USD	0,090 USD	0,080 USD

Criar uma Distribuição

CloudFront Origin Access Identity (OAI): ao se usar OAI, buckets podem restringir acesso ao conteúdo através do CloudFront, caso não seja OAI, o acesso ao bucket deve ser público

Origin Shield (Escudo de Origem): é um caching adicional que pode ajudar a reduzir o carregamento sobre a sua origem e ajudar a proteger sua disponibilidade

SSL Certificado: é possível associar um certificado do AWS ACM para conexões HTTPS

AWS WAF: O AWS WAF é um firewall de aplicações Web que ajuda a proteger suas aplicações Web ou APIs contra bots e exploits comuns na Web que podem afetar a disponibilidade, comprometer a segurança ou consumir recursos em excesso.

Origens: Após a criação da distribuição, é possível adicionar mais de uma origem de fonte de dados que serão distribuídos para as Edge Locations

Behavior: é possível alterar o comportamento de cada origem selecionada, como por exemplo o protocolo de acesso (HTTP/HTTPS), os métodos HTTP, política de cache e política da solicitação de origem, além de alterar as funções associadas na request e response da origem e do visualizador

URL Assinadas: Um signed URL inclui informações adicionais, por exemplo, uma data e hora de expiração, que proporcionam a você mais controle sobre o acesso a seu conteúdo. Suponha que outros sites estão linkando a mesma imagem (armazenada no S3) que está disponível no seu site gerando custos adicionais pelos acessos. A solução é utilizar URLs Assinadas com data de expiração pois assim, de tempos em tempos, a URL mudará e os acessantes parasitas perderão o acesso aquele link após a expiração da URL. Exemplo de URL Assinada:

```
❶ http://d111111abcdef8.cloudfront.net/image.jpg ❷ ? ❸ color=red&size=medium& ❹
Expires=1357034400 ❺ &Signature=nitfHRCrtziwO2HwPfWw~yYDhUF5EwRunQA-
j19DzZrvDh6hQ73lDx~-ar3UocvvRQVw6EkC~GdpGQyyOSKQim-TxAnW7d8F5Kkai9HVx0FIu-
5jcQb0UEmatEXAMPLE3ReXySpLSMj0yCd3ZAB4UcBCAqEijkytL6f3fVYNGQI6 ❻ &Key-Pair-
Id=K2JCJMDEHXQW5F
```

Error Pages: Quando sua origem retorna um erro, você pode escolher por quanto tempo (TTL) o CloudFront armazena em cache o erro. Você também pode optar por enviar uma resposta personalizada ao visualizador com um código de resposta HTTP diferente, em vez de encaminhar o erro que o CloudFront recebeu da origem. Obviamente que uma página customizada pode ser setada na visualização do erro

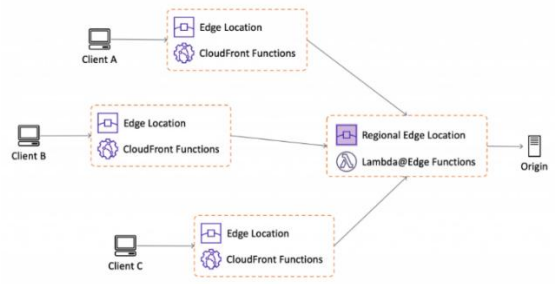
Geographic Restrictions: é possível criar Block list ou Allow list de países que não terão ou terão acesso ao recurso pelo CloudFront

Uma distribuição habilitada não pode ser excluída, apenas desabilitada: Quando está desativado, a distribuição fica offline e não pode responder às solicitações. Você pode habilitar a distribuição posteriormente para restaurá-la.

CloudFront Functions

Escreva funções de ponta leves em JavaScript usando funções do CloudFront: Com o CloudFront Functions no Amazon CloudFront, você pode escrever funções leves em JavaScript para personalizações CDN de alta escala e sensíveis à latência. Suas funções podem manipular as solicitações e respostas que fluem pelo CloudFront, executar autenticação e autorização básicas, gerar respostas HTTP na extremidade e muito mais. O ambiente de tempo de execução do CloudFront Functions oferece tempos de inicialização de submilissegundos, dimensiona imediatamente para lidar com milhões de solicitações por segundo e é altamente seguro. CloudFront Functions é um recurso nativo do CloudFront, o que significa que você pode criar, testar e implantar seu código inteiramente dentro do CloudFront

Function Associations: Com funções de ponta, você pode escrever seu próprio código para personalizar como sua distribuição do CloudFront processa solicitações e respostas HTTP. Você pode ter até quatro funções de borda por comportamento de cache, uma para cada tipo de evento: solicitação do visualizador, resposta do visualizador, solicitação de origem e resposta de origem. Você pode escolher entre dois tipos de funções de borda, CloudFront Functions e Lambda@Edge. Funções do CloudFront estão disponíveis apenas para solicitações do visualizador e tipos de eventos de resposta do visualizador. Com Lambda@Edge, seu código de função pode acessar o corpo da solicitação HTTP.

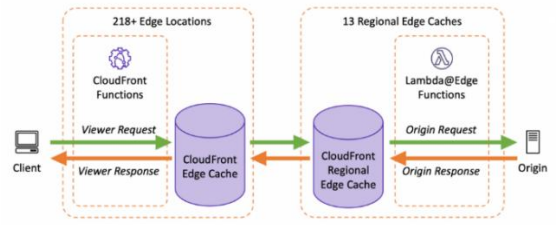


<u>Funções do CloudFront são ideais para processamento leve de solicitações da web, por exemplo</u>:

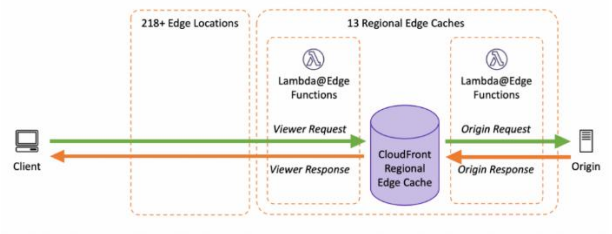
Manipulações e normalização da chave do cache: Transforme os atributos de solicitação HTTP (como URL, cabeçalhos, cookies e strings de consulta) para construir a chave do cache, que é o identificador exclusivo para objetos no cache e é usado para determinar se um objeto já existe em cache.

Reescritas e redirecionamentos de URL: Gere uma resposta para redirecionar solicitações para uma URL diferente. Por exemplo, redirecione um usuário não autenticado de uma página restrita para um formulário de login. As regravações de URL também podem ser usadas para testes A/B (*O teste A / B é uma maneira de comparar duas versões de uma única variável, normalmente testando a resposta de um sujeito à variante A contra a variante B e determinando qual das duas variantes é mais eficaz - Wikipedia*)

Manipulação de cabeçalho HTTP: Visualize, adicione, modifique ou exclua qualquer um dos cabeçalhos de solicitação/resposta. Por exemplo, adicione cabeçalhos HTTP Strict Transport Security (HSTS) à sua resposta ou copie o endereço IP do cliente em um novo cabeçalho HTTP para que seja encaminhado para a origem com a solicitação.



Se você precisar de alguns dos recursos do Lambda@Edge que não estão disponíveis com o CloudFront Functions, como acesso à rede ou um tempo de execução mais longo, você ainda pode usar o Lambda @ Edge antes e depois do conteúdo ser armazenado em cache pelo CloudFront.



Updating Existing Content with a CloudFront Distribution

There are two ways to update existing content that CloudFront is set up to distribute for you:

- Update files by using the same name
- Update by using a version identifier in the file name

We recommend that you use a version identifier in file names or in folder names, to help give you more control over managing the content that CloudFront serves.

Updating Existing Files Using Versioned File Names

When you update existing files in a CloudFront distribution, we recommend that you include some sort of version identifier either in your file names or in your directory names to give yourself better control over your content. This identifier might be a date-time stamp, a sequential number, or some other method of distinguishing two versions of the same object.

For example, instead of naming a graphic file image.jpg, you might call it image_1.jpg. When you want to start serving a new version of the file, you'd name the new file image_2.jpg, and you'd update the links in your web application or website to point to image_2.jpg. Alternatively, you might put all graphics in an images_v1 directory and, when you want to start serving new versions of one or more graphics, you'd create a new images_v2 directory, and you'd update your links to point to that directory. With versioning, you don't have to wait for an object to expire before CloudFront begins to serve a new version of it, and you don't have to pay for object invalidation.

Even if you version your files, we still recommend that you set an expiration date.

Updating Existing Content Using the Same File Names

Although you can update existing files in a CloudFront distribution and use the same file names, we don't recommend it. CloudFront distributes files to edge locations only when the files are requested, not when you put new or updated files in your origin. If you update an existing file in your origin with a newer version that has the same name, an edge location won't get that new version from your origin until both of the following occur:

- The old version of the file in the cache expires.
- There's a user request for the file at that edge location.

Sistema de Cobrança CloudFront

CloudFront is a content delivery network (CDN) service that offers a simple, pay-as-you-go pricing model

<u>Serviços Cobrados</u>

- Data transfer out to the Internet from edge locations.
- Data transfer out of Amazon CloudFront to the origin server
- Custom SSL certificate associated with the CloudFront distribution using the Dedicated IP version of custom SSL certificate support. Because for each custom SSL certificate associated with one or more CloudFront distributions using the Dedicated IP version of custom SSL certificate support, you are charged at $600 per month.

<u>Serviços Gratuitos</u>

- Data transfer from origin to CloudFront edge locations (Amazon CloudFront "origin fetches")

Managing how long content stays in the cache (expiration)

You can control how long your files stay in a CloudFront cache before CloudFront forwards another request to your origin. Reducing the duration allows you to serve **dynamic content**. Increasing the duration means that your users get better performance because your files are more likely to be served directly from the edge cache. A longer duration also reduces the load on your origin.

Typically, CloudFront serves a file from an edge location until the cache duration that you specified passes—that is, until the file expires. After it expires, the next time the edge location gets a user request for the file, CloudFront forwards the request to the origin server to verify that the cache contains the latest version of the file. The response from the origin depends on whether the file has changed:

If the CloudFront cache already has the latest version, the origin returns a status code 304 Not Modified.

If the CloudFront cache does not have the latest version, the origin returns a status code 200 OK and the latest version of the file.

If a file in an edge location isn't frequently requested, CloudFront might evict the file—remove the file before its expiration date—to make room for files that have been requested more recently.

By default, each file automatically expires after 24 hours, but you can change the default behavior in two ways:

- **To change the cache duration for all files** that match the same path pattern, you can change the CloudFront settings for Minimum TTL, Maximum TTL, and Default TTL for a cache behavior. For information about the individual settings, see Minimum TTL, Maximum TTL, and Default TTL in Values That You Specify When You Create or Update a Distribution. To use these settings, you must choose the Customize option for the Object Caching setting when you create or update your CloudFront distribution. For more information, see Object Caching in Values That You Specify When You Create or Update a Distribution.

- **To change the cache duration for an individual file**, you can configure your origin to add a Cache-Control max-age or Cache-Control s-maxage directive, or an Expires header field to the file. For more information, see Using headers to control cache duration for individual objects.

Links Úteis

https://aws.amazon.com/pt/blogs/aws/introducing-cloudfront-functions-run-your-code-at-the-edge-with-low-latency-at-any-scale/

https://docs.aws.amazon.com/pt_br/AmazonCloudFront/latest/DeveloperGuide/private-content-creating-signed-url-canned-policy.html

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/Expiration.html

➢ Pontos de Atenção

1. Um aplicativo da web está usando o CloudFront para distribuir suas imagens, vídeos e outros conteúdos estáticos armazenados em seu balde S3 para seus usuários em todo o mundo. A empresa lançou recentemente um novo acesso exclusivo para membros a alguns de seus arquivos de mídia de alta qualidade. Há um requisito para fornecer acesso a vários arquivos de mídia privados apenas para seus assinantes pagantes, sem ter que alterar seus URLs atuais. Qual das alternativas a seguir é a solução mais adequada que você deve implementar para atender a esse requisito? R: Use cookies assinados para controlar quem pode acessar os arquivos privados em sua distribuição do CloudFront, modificando seu aplicativo para determinar se um usuário deve ter acesso ao seu conteúdo. Para membros, envie os header Set-Cookie necessários para o visualizador, que desbloqueará o conteúdo apenas para eles.

2. A global medical research company has a molecular imaging system which provides each client with frequently updated images of what is happening inside the human body at the molecular and cellular level. The system is hosted in AWS and the images are hosted in an S3 bucket behind a CloudFront web distribution. There was a new batch of updated images that were uploaded in S3, however, the users were reporting that they were still seeing the old content. You need to control which image will be returned by the system even when the user has another version cached either locally or behind a corporate caching proxy. Which of the following is the most suitable solution to solve this issue? **R: Use versioned objects**

3. A web application is hosted in AWS EC2 and serves global customers. As the application is getting more and more popular, the data transfer cost keeps increasing. You plan to use AWS CloudFront to improve the latency and reduce the cost. Which of the following services is free for CloudFront? **R: Data transfer from origin to CloudFront edge locations (Amazon CloudFront "origin fetches").**

4. Your application is deployed in EC2 instances and uses CloudFront to deliver the content. In order to reduce the cost of requests to the origin, you plan to increase the cache duration in CloudFront for certain dynamic contents. Which of the following options is the most appropriate to achieve the requirement? R: Modify the application to add a Cache-Control header to control how long the objects stay in the CloudFront cache.

5. Você está desenvolvendo um site para uma empresa que transmite vídeos de anime. Você serve esse conteúdo por meio do CloudFront. A empresa implementou uma seção para assinantes premium. Esta seção contém mais vídeos do que a seção gratuita. Você deseja garantir que apenas assinantes premium possam acessar esta seção premium. Como você pode conseguir isso facilmente? R: Using CloudFront origin with signed cookies. With a signed URL, you only restrict access to one file. If you implement this, the application would have to generate a new pre-signed URL for every file in the premium section and give this new link to the user. Once your application allows your user to see this section, it will give a cookie to the user. Then CloudFront will use this cookie to access S3 restricted content.

6. Você configura um bucket do Amazon S3 como a origem para uma nova distribuição do CloudFront. Você precisa restringir o acesso para não visualizar os arquivos usando diretamente os URLs do S3. Os arquivos só devem ser buscados por meio do URL do CloudFront. Qual método é o mais adequado? R: Create the origin access identity (OAI) and associate it with the distribution. Because you can configure the CloudFront origin to restrict bucket access through OAI.

7. Você tem um aplicativo da web hospedado em uma instância EC2 na AWS que os usuários podem acessar em todo o mundo. A equipe de Operações tem recebido solicitações de suporte sobre lentidão extrema de usuários em algumas regiões. O que pode ser feito com a arquitetura para melhorar o tempo de resposta para esses usuários? R: Place the EC2 Instance behind CloudFront.

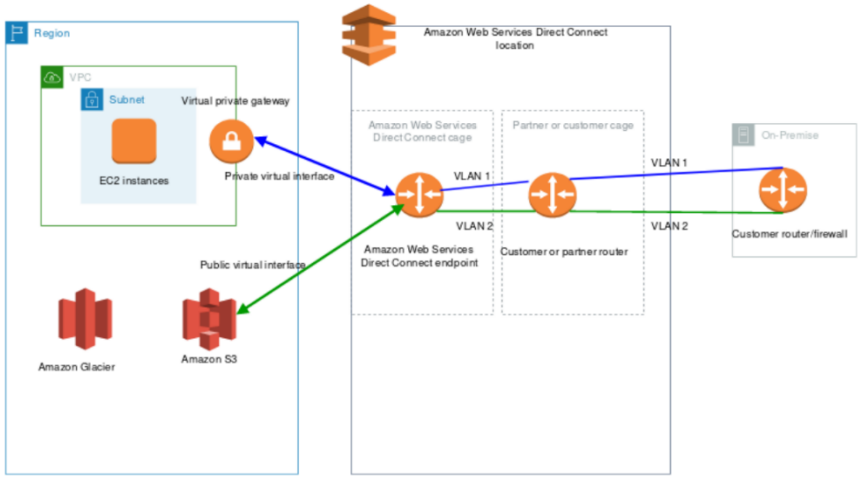
8. Uma grande empresa de TI está usando o Amazon CloudFront para seu aplicativo da web. O conteúdo estático para este aplicativo é salvo no bucket do Amazon S3. O Amazon CloudFront está configurado para este aplicativo para fornecer acesso mais rápido a esses arquivos para usuários globais. A equipe de TI está preocupada com alguns arquivos críticos que precisam ser acessados apenas por usuários de determinados pools de IP de lista branca que você definiu no Amazon CloudFront e nenhum usuário deve acessar esses arquivos diretamente usando o URL do Amazon S3. Qual das opções a seguir é a maneira mais segura de controlar o acesso a esses arquivos? R: Create an OAI user to associate with distribution & modify permission on Amazon S3 bucket using object ACL's. Amazon CloudFront Origin Access Identity is a special user that can control access to content in the Amazon S3 bucket. Using Object ACLs provides a granular control on each file in the Amazon S3 bucket. Associating Amazon CloudFront OAI to a distribution & modifying permission on the S3 bucket allows access only to OAI. It ensures that no users can directly access content in the S3 bucket, & all access passes through Amazon CloudFront using OAI.

AWS Direct Connect

O AWS Direct Connect conecta sua rede interna a um local do AWS Direct Connect por meio de um cabo de fibra óptica Ethernet padrão. Uma extremidade do cabo é conectada ao seu roteador, a outra a um roteador AWS Direct Connect.

Com essa conexão, você pode criar interfaces virtuais diretamente para serviços públicos da AWS (por exemplo, para Amazon S3) ou para Amazon VPC, ignorando provedores de serviços de Internet em seu caminho de rede. Um local do AWS Direct Connect fornece acesso ao AWS na região à qual está associado.

Você pode usar uma única conexão em uma região pública ou AWSGovCloud (EUA) para acessar serviços públicos da AWS em todas as outras regiões públicas. O diagrama a seguir mostra como o AWS Direct Connect faz interface com sua rede.



Trocando em Miúdos

O Direct Connect funciona em conjunto com as operadoras de Telecom, que em conjunto com a AWS cria um link dedicado (*Leased Line - MPLS). Normalmente são instalados roteadores da Telecom dentro da Infra da AWS. O Link dedicado respeita o valor de download e upload adquiridos. Depois que um Direct Connect é criado a topologia de rede da empresa terá acesso direto a VPC

Leased Line (linha dedicada): Uma linha alugada é um circuito privado de telecomunicações entre dois ou mais locais fornecidos de acordo com um contrato comercial. Às vezes também é conhecido como circuito privado

AWS Direct Connect Components

Connection: Crie uma conexão em um local do AWS Direct Connect para estabelecer uma conexão de rede de suas instalações para uma região da AWS.

Virtual interfaces: Crie uma interface virtual para permitir o acesso aos serviços da AWS. Uma interface virtual pública permite o acesso a serviços públicos, como Amazon S3. Uma interface virtual privada permite o acesso ao seu VPC.

Requisitos de Network

Para usar o AWS Direct Connect em um local do AWS Direct Connect, sua rede deve atender a uma das seguintes condições:

- Sua rede está colocada em um local existente do AWS Direct Connect.
- Você está trabalhando com um parceiro AWS Direct Connect que é membro da AWS Partner Network (APN).
- Você está trabalhando com um provedor de serviços independente para se conectar ao AWS Direct Connect.

AWS Global Accelerator

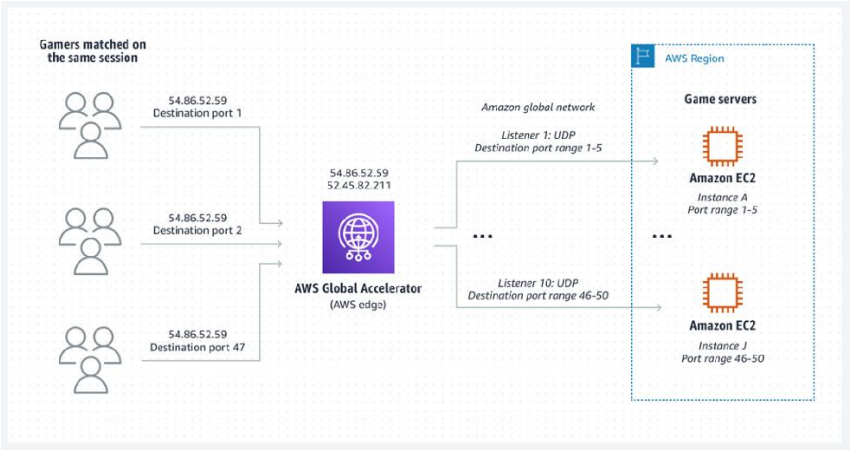
O AWS Global Accelerator é um serviço de rede que melhora o desempenho do tráfego de seus usuários em até **60%** usando a infraestrutura de rede global da Amazon Web Services (Edge Locations).

Quando a Internet está congestionada, o AWS Global Accelerator otimiza o caminho para seu aplicativo para manter a perda de pacotes, o jitter e a latência consistentemente baixos.

Com o Global Accelerator, você recebe dois IPs públicos estáticos globais que atuam como um ponto de entrada fixo para sua aplicação, melhorando a disponibilidade.

No back-end, adicione ou remova os endpoints das suas aplicações da AWS, como os Application Load Balancers, Network Load Balancers, instâncias do EC2 e IPs elásticos, sem fazer alterações voltadas para o usuário.

O Global Accelerator redireciona automaticamente seu tráfego para o endpoint disponível mais próximo para mitigar falhas de endpoints.



➢ Pontos de Atenção

Você tem um aplicativo hospedado no EC2. Este aplicativo é acessado por usuários em diferentes países e precisa de alto desempenho. O aplicativo também precisa de IPs estáticos que não devem ser alterados. Qual serviço pode ser usado para atender a esses requisitos? R: AWS Global Accelerator. *AWS Global Accelerator provides two static IPs as endpoints. AWS Global Accelerator also improves the application's performance by directing traffic to endpoints over the AWS global network.*

Amazon Route 53

O Amazon Route 53 é um web service Domain Name System (DNS) na nuvem altamente disponível e escalável

Ele foi projetado para oferecer aos desenvolvedores e empresas uma maneira altamente confiável e econômica de direcionar aos usuários finais aos aplicativos de internet

Conecta com eficiência as solicitações de usuários com a infraestrutura executada na AWS, como instâncias do Amazon EC2, load balancers do Elastic Load Balancing e buckets do Amazon S3 e também pode ser usado para rotear usuários para infraestruturas fora da AWS

O fluxo de tráfego do Amazon Route 53 facilita o gerenciamento de tráfego de maneira global por meio de diversos tipos de roteamento: **Latency Based Routing, Geo DNS, geoproximidade e Weighted Round Robin**

Todos os tipos de roteamento podem ser combinados com o **failover de DNS**, viabilizando diversas arquiteturas de baixa latência e tolerantes a falhas

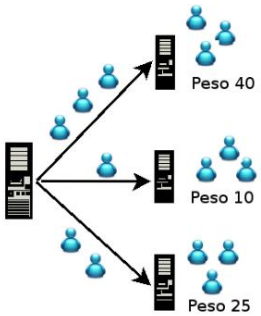
O Amazon Route 53 também oferece registro de nome de domínio, ou seja, você pode comprar e gerenciar nomes de domínio (como exemplo.com) e o Route 53 configura automaticamente as definições de DNS para os seus domínios

Observação: domínios que sejam .com.br não estão disponíveis para serem adquiridos pelo Route 53, caso se deseje um domínio com .br será necessário adquirir externamente em outra plataforma, como Registro.br por exemplo

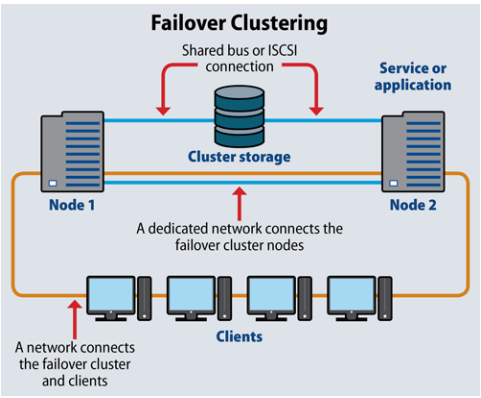
Latency Based Routing (Roteamento baseado em latência): Roteie usuários finais para a região da AWS que fornece a latência mais baixa possível, pois o Route 53 tem arquitetura global

Geo DNS: Roteie usuários finais a um determinado endpoint especificado de acordo com a localização geográfica do usuário final. Exemplo: existe dois buckets, se o usuário acessar dos EUA vai para o bucket A e se acessar do Brasil vai para o bucket B

Weighted Round Robin: O Amazon Route 53 oferece a funcionalidade Weighted Round Robin (WRR), permitindo especificar a frequência (“pesos”) com que diferentes respostas DNS são retornadas aos usuários finais. É utilizado para realizar testes A/B ou balancear o trafego entre regiões ou data centers de tamanhos variados



Failover de DNS: Encaminhe automaticamente os visitantes do seu site a um local alternativo para evitar paralizações do site, ou seja, cria um fluxo alternativo para outro local caso o primeiro tenha tido algum tipo de falha



Características

Responsável pelas resoluções de endereçamento IP

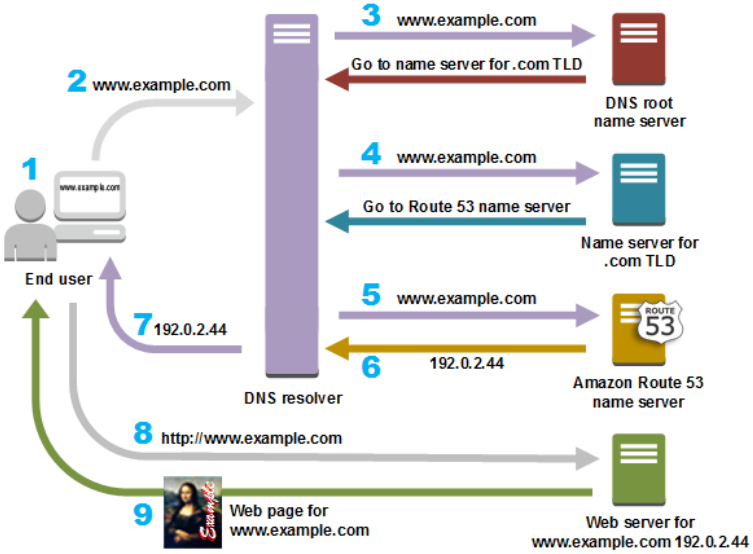
Redundância de todas as localizações

100% de SLA (Availability)

Geo Location

Failover

Como Funciona o DNS



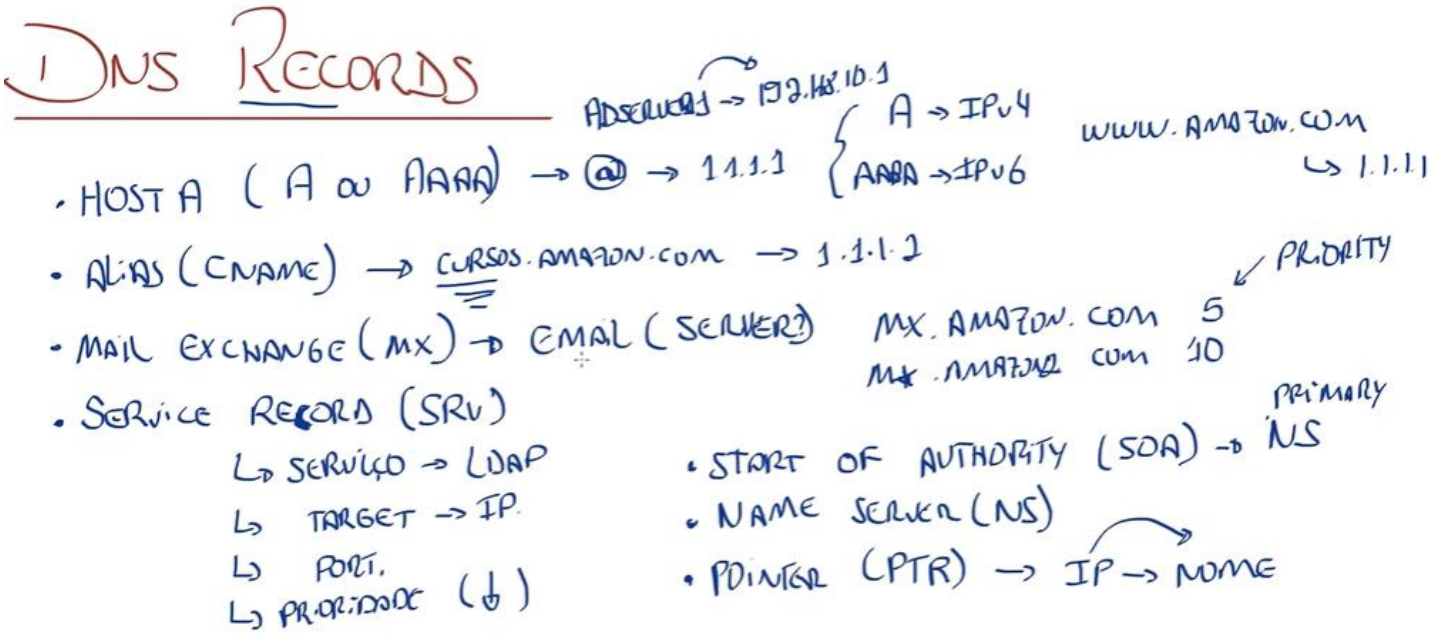
Tipos de Registro de DNS

Host A (A ou AAAA) – www.amazon.com / 1.1.1.1 – A/IPv4 e AAAA/IPv6

Alias (Cname) – cursos.amazon.com / 1.1.1.2

Mail Exchange (MX) – email (server?) – mx.amazon.com 5 / mx.amazon.com 10 (5 e 10 é a prioridade)

Service Record (SRV) –



Quais tipos de registro DNS são compatíveis com o Amazon Route 53?

No momento, o Amazon Route 53 é compatível com os seguintes tipos de registro DNS:

- A (address record – registro de endereço)
- AAAA (IPv6 address record – registro de endereço IPv6)
- CNAME (canonical name record – registro de nome canônico)
- Certification authority authorization (CAA – Autorização de autoridade de certificação)
- Mail exchange record (MX – Registro de troca de e-mail)
- NAPTR (registro de Name Authority Pointer)
- NS (name server record – registro do servidor do nome)
- PTR (pointer record – registro do apontador)
- SOA (start of authority record – registro de início da autoridade)
- SPF (sender policy framework – estrutura da política do remetente)
- SRV (service locator – localizador do serviço)
- TXT (registro de texto)

O Amazon Route 53 também oferece registros de alias, que são uma extensão específica do Amazon Route 53 para DNS. É possível criar registros de alias para rotear o tráfego para recursos da AWS selecionados, incluindo load balancers do Amazon Elastic Load Balancing, distribuições do Amazon CloudFront, ambientes do AWS Elastic Beanstalk, Gateways de API, endpoints da interface da VPC e buckets do Amazon S3 configurados como websites. O registro de alias geralmente tem um tipo de A ou AAAA, mas funciona como um registro CNAME. Usando um registro de alias, você pode mapear o nome do seu registro (example.com) para o nome DNS de um recurso da AWS (elb1234.elb.amazonaws.com). Os resolvedores veem o registro A ou AAAA e o endereço IP do recurso da AWS.

O registro CNAME mapeia um nome para outro nome. Só deve ser usado quando não houver outros registros com esse nome. O registro ALIAS mapeia um nome para outro nome, mas pode coexistir com outros registros naquele nome.

Resources that you can redirect queries to

Alias records

An alias record can only redirect queries to selected AWS resources, such as the following:

- Amazon S3 buckets
- CloudFront distributions
- Another record in the same Route 53 hosted zone

For example, you can create an alias record named acme.example.com that redirects queries to an Amazon S3 bucket that is also named acme.example.com. You can also create an acme.example.com alias record that redirects queries to a record named zenith.example.com in the example.com hosted zone.

CNAME records

A CNAME record can redirect DNS queries to any DNS record. For example, you can create a CNAME record that redirects queries from acme.example.com to zenith.example.com or to acme.example.org. You don't need to use Route 53 as the DNS service for the domain that you're redirecting queries to.

Resumo

While ordinary Amazon Route 53 records are standard DNS records, *alias records* provide a Route 53–specific extension to the DNS functionality. Instead of an IP address or a domain name, an alias record contains a pointer to a

CloudFront distribution, an Elastic Beanstalk environment, an ELB Classic, Application, or Network Load Balancer, an Amazon S3 bucket that is configured as a static website, or another Route 53 record in the same hosted zone. When Route 53 receives a DNS query that matches the name and type in an alias record, Route 53 follows the pointer and responds with the applicable value.

Note: Route 53 uses "Alias Name" to connect to the CloudFront as Alias Record is a Route 53 extension to DNS. Also, an Alias record is similar to a CNAME record, but the main difference is - you can create an Alias record for both root domain & subdomain. In contrast, a CNAME record can be created only to subdomain.

Trocando em Miúdos: Host, CNAME e Alias

var.example.com. CNAME foo.example.com.

bar.example.com. CNAME foo.example.com.

foo.example.com. A 192.0.2.23

AWS ROUTE53 ALIAS-

Alias resource record sets provide an Amazon Route 53–specific extension to DNS functionality.

Alias Usecase

ELB`s, CDNs não têm endereços IPV4 predefinidos, você deve resolvê-los usando um nome de DNS, portanto, se você precisar criar um registro de apex de zona (por exemplo, exemplo.com, acloud.guru), você precisará criar usando Os registros ALIAS permitem que você resolva seu nome de domínio sem www para endereço ELB ou CDN DNS, mapeando o nome de domínio sem www com o nome de ELB / CDN DNS.

Pontos a serem observados - se um conjunto de registros de recurso de alias aponta para uma distribuição do CloudFront, um ambiente Elastic Beanstalk, um balanceador de carga ELB ou um bucket do Amazon S3, você não pode definir o tempo de vida (TTL); O Amazon Route 53 usa CloudFront, Elastic Beanstalk, Elastic Load Balancing ou Amazon S3 TTLs.

Exemplo:

1 - Your organization had created an S3 bucket for static website hosting. They had created and configured all necessary components for the static website with the S3 website endpoint http://example-bucket.com.s3-website-us-east-2.amazonaws.com. They would like to get the website served through the domain name www.example-bucket.com which is already registered. You want to create a record in Route 53 to route to the S3 website endpoint. Which type of record set you need to create?

A. A – IPv4 Address with Alias=NO
B. A – IPv4 Address with Alias=YES
C. CNAME – Canonical Name with ALIAS=NO
D. CNAME – Canonical Name with ALIAS=YES

R: B. A – IPv4 Address with Alias=YES

We can't create a CNAME record for the Parent, Naked, or Apex domain, whereas we can use an alias record to point to the parent domain.
Below scenarios given are not possible.
www.example-bucket.com CNAME http://example-bucket.com.s3-website-us-east-2.amazonaws.com
or
example-bucket.com CNAME http://example-bucket.com.s3-website-us-east-2.amazonaws.com
But with an alias, they are possible.
example-bucket.com Alias(A) http://example-bucket.com.s3-website-us-east-2.amazonaws.com

2 - You have launched an RDS instance in your VPC. The endpoint that is assigned to your DB instance is a long, partially random, alphanumeric string, for example, myexampledb.a1b2c3d4wxyz.us-west-2.rds.amazonaws.com.

Your organization wants to use a name that's easier to remember. So you registered a domain name using the Route53 service. Which type of record set do you need to create?

R: CNAME - Canonical Name with ALIAS=NO

Note:
A Record (Address Record) is typically used when an IP address to name conversion is required. This is most commonly used.

A record points a name to specific IP. If you want whizlabs.com to point to the server 10.120.13.14, you will configure an A record.

whizlabs.com A 10.120.13.14

A CNAME record (Canonical record) points a name to another name. Typically a complex name with alphanumeric characters can be shortened to a user understandable format.

blog.whizlabs.com CNAME blog235_github_repoABCD.io.net

In the question, the AWS generated resource name is complex i.e myexambledb.a1b2c3d4.....

This can be converted to a user-friendly name by using a CNAME record.

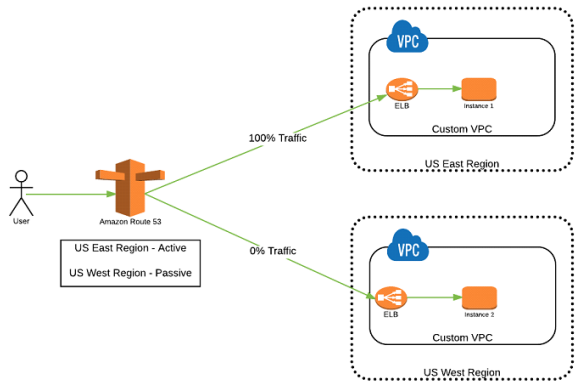
Métricas

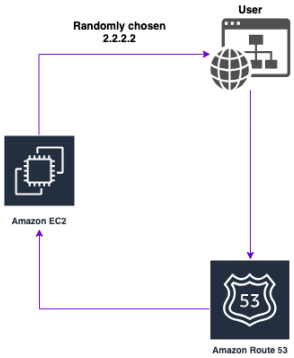
Choosing a Routing Policy

When you create a record, you choose a routing policy, which determines how Amazon Route 53 responds to queries:

- **Simple routing policy** – Use for a single resource that performs a given function for your domain, for example, a web server that serves content for the example.com website.
- **Failover routing policy** – Use when you want to configure active-passive failover.
- **Geolocation routing policy** – Use when you want to route traffic based on the location of your users.
- **Geoproximity routing policy** – Use when you want to route traffic based on the location of your resources and, optionally, shift traffic from resources in one location to resources in another.
- **Latency routing policy** – Use when you have resources in multiple AWS Regions and you want to route traffic to the region that provides the best latency.
- **Multivalue answer routing policy** – Use when you want Route 53 to respond to DNS queries with up to eight healthy records selected at random.
- **Weighted routing policy** – Use to route traffic to multiple resources in proportions that you specify.

Simple Routing





Returned Values
1.1.1.1
2.2.2.2
3.3.3.3

Failover Routing

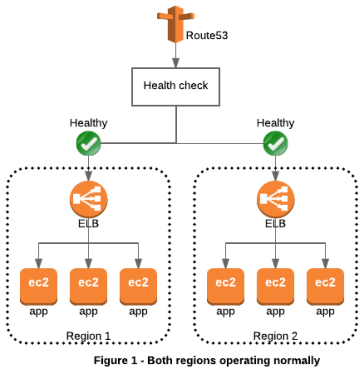


Figure 1 - Both regions operating normally

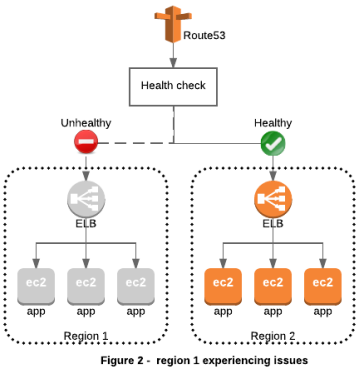
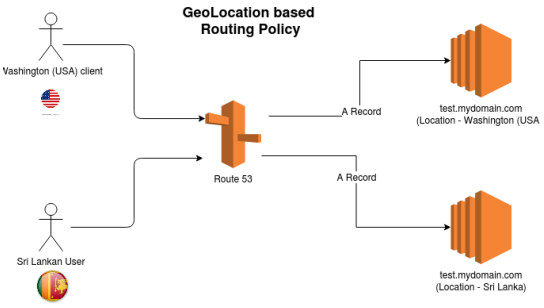
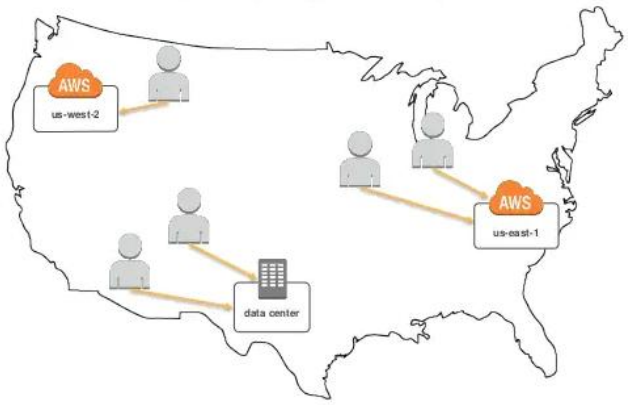


Figure 2 - region 1 experiencing issues

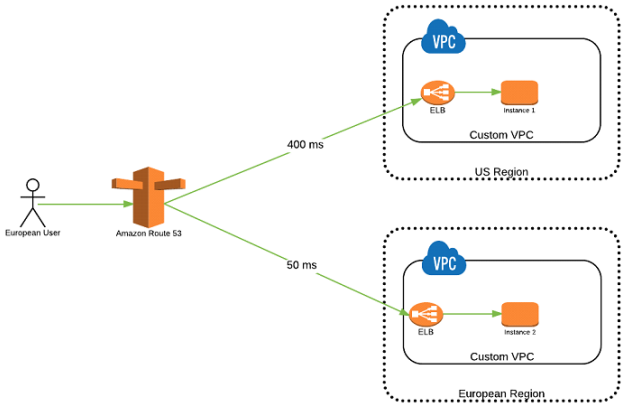
Geolocation Routing



Geoproximity Routing (Traffic Flow Only)



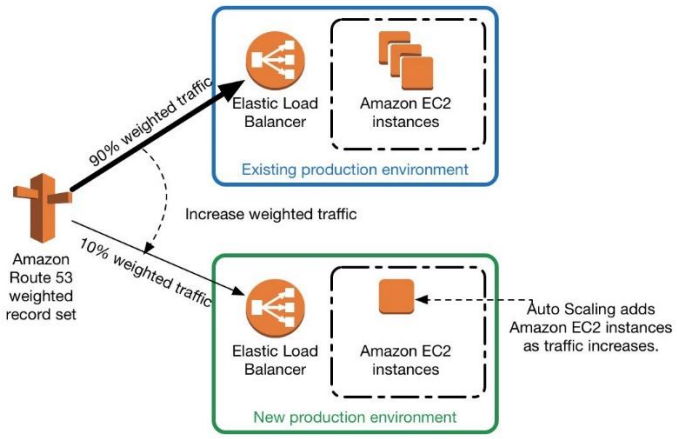
Latency-based Routing



Multivalue Answer Routing

É o mesmo que Simple Routing, faz o roteamento de forma randômica, porém com Health Check

Weighted Routing



Active-active and active-passive failover

You can use Route 53 health checking to configure active-active and active-passive failover configurations. You configure active-active failover using any routing policy (or combination of routing policies) other than failover, and you configure active-passive failover using the failover routing policy.

Active-active failover

Use this failover configuration when you want all of your resources to be available the majority of the time. When a resource becomes unavailable, Route 53 can detect that it's unhealthy and stop including it when responding to queries.

In active-active failover, all the records that have the same name, the same type (such as A or AAAA), and the same routing policy (such as weighted or latency) are active unless Route 53 considers them unhealthy. Route 53 can respond to a DNS query using any healthy record.

Active-passive failover

Use an active-passive failover configuration when you want a primary resource or group of resources to be available the majority of the time and you want a secondary resource or group of resources to be on standby in case all the

primary resources become unavailable. When responding to queries, Route 53 includes only the healthy primary resources. If all the primary resources are unhealthy, Route 53 begins to include only the healthy secondary resources in response to DNS queries.

Configuring DNSSEC for a domain

Às vezes, os invasores sequestram o tráfego para endpoints da Internet, como servidores da Web, interceptando consultas DNS e retornando seus próprios endereços IP aos resolvedores DNS no lugar dos endereços IP reais desses endpoints. Os usuários são então roteados para os endereços IP fornecidos pelos invasores na resposta falsificada, por exemplo, para sites falsos. Você pode proteger seu domínio desse tipo de ataque, conhecido como falsificação de DNS ou ataque man-in-the-middle, configurando Extensões de Segurança do Sistema de Nomes de Domínio (DNSSEC), um protocolo para proteger o tráfego DNS.

Amazon Route 53 supports DNSSEC signing as well as DNSSEC for domain registration. If you want to configure DNSSEC signing for a domain that's registered with Route 53

Weighted Records

Managing Over 100 Weighted Records in Amazon Route 53

Amazon Route 53 lets you configure weighted records. For a given name and type (for example, `www.example.com`, type A), you can configure up to 100 alternative responses, each with its own weight. When responding to queries for `www.example.com`, Route 53 DNS servers select a weighted random response to return to DNS resolvers. The value of a weighted record that has a weight of 2 is returned, on average, twice as often as the value of a weighted record that has a weight of 1.

If you need to direct traffic to more than 100 endpoints, one way to achieve this is to use a tree of weighted alias records and weighted records. For example, the first "level" of the tree may be up to 100 weighted alias records, each of which can, in turn, point to up to 100 weighted records. Route 53 permits up to three levels of recursion, allowing you to manage up to 1,000,000 unique weighted endpoints.

A simple two-level tree might look like this:

Weighted alias records

- `www.example.com` aliases to `www-a.example.com` with a weight of 1
- `www.example.com` aliases to `www-b.example.com` with a weight of 1

Weighted records

- `www-a.example.com`, type A, value 192.0.2.1, weight 1
- `www-a.example.com`, type A, value 192.0.2.2, weight 1

- `www-b.example.com`, type A, value 192.0.2.3, weight 1
- `www-b.example.com`, type A, value 192.0.2.4, weight 1

Route 53 Health Checks

Types of Amazon Route 53 Health Checks

You can create three types of Amazon Route 53 health checks:

Health checks that monitor an endpoint

You can configure a health check that monitors an endpoint that you specify either by IP address or by domain name. At regular intervals that you specify, Route 53 submits automated requests over the internet to your application, server, or other resource to verify that it's reachable, available, and functional. Optionally, you can configure the health check to make requests similar to those that your users make, such as requesting a web page from a specific URL.

Health checks that monitor other health checks (calculated health checks)

You can create a health check that monitors whether Route 53 considers other health checks healthy or unhealthy. One situation where this might be useful is when you have multiple resources that perform the same function, such as multiple web servers, and your chief concern is whether some minimum number of your resources are healthy. You can create a health check for each resource without configuring notification for those health checks. Then you can create a health check that monitors the status of the other health checks and that notifies you only when the number of available web resources drops below a specified threshold.

Health checks that monitor CloudWatch alarms

You can create CloudWatch alarms that monitor the status of CloudWatch metrics, such as the number of throttled read events for an Amazon DynamoDB database or the number of Elastic Load Balancing hosts that are considered healthy. After you create an alarm, you can create a health check that monitors the same data stream that CloudWatch monitors for the alarm.

To improve resiliency and availability, Route 53 doesn't wait for the CloudWatch alarm to go into the `ALARM` state. The status of a health check changes from healthy to unhealthy based on the data stream and on the criteria in the CloudWatch alarm.

Definição de Preço

- Gerenciamento de zonas hospedadas: Você paga um encargo mensal para cada zona hospedada gerenciada com o Route 53.
- Atendimento a consultas do DNS: Haverá cobrança para cada consulta do DNS respondida pelo serviço Amazon Route 53, exceto para consultas a registros Alias A mapeadas para instâncias do Elastic Load Balancing, distribuições do CloudFront, ambientes do AWS Elastic Beanstalk, API Gateways, VPC endpoints ou buckets do site do Amazon S3, fornecidas gratuitamente.
- Gerenciamento de nomes de domínios: Você paga uma taxa anual para cada nome de domínio registrado pelo ou transferido para o Route 53.

Zonas hospedadas e registros

- 0,50 USD por zona hospedada/mês para as primeiras 25 zonas hospedadas
- 0,10 USD por zona hospedada/mês para zonas hospedadas adicionais

Consultas padrão

- 0,40 USD por milhão de consultas – primeiro 1 bilhão de consultas/mês
- 0,20 USD por milhão de consultas – mais de 1 bilhão de consultas/mês

Consultas de roteamento baseado em latência

- 0,60 USD por milhão de consultas – primeiro 1 bilhão de consultas/mês
- 0,30 USD por milhão de consultas – mais de 1 bilhão de consultas/mês

Consultas de Geo DNS e geoproximidade

- 0,70 USD por milhão de consultas -- primeiro 1 bilhão de consultas/mês
- 0,35 USD por milhão de consultas -- mais de 1 bilhão de consultas/mês

Links Úteis

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/TutorialAddingLBRRegion.html

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html#routing-policy-latency

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resource-record-sets-choosing-alias-non-alias.html

https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-to-cloudfront-distribution.html

➢ Pontos de Atenção

1. You are setting up the cloud architecture for an international money transfer service to be deployed in AWS which will have thousands of users around the globe. The service should be available 24/7 to avoid any business disruption and should be resilient enough to handle the outage of an entire AWS region. To meet this requirement, you have deployed your AWS resources to multiple AWS Regions. You need to use Route 53 and configure it to set all of your resources to be available all the time as much as possible. When a resource becomes unavailable, your Route 53 should detect that it's unhealthy and stop including it when responding to queries. Which of the following is the most fault tolerant routing configuration that you should use in this scenario? R: Configure an Active-Active Failover with Weighted routing policy.

2. You are working as the Solutions Architect for a global technology consultancy firm which has an application that uses multiple EC2 instances located in various AWS regions such as US East (Ohio), US West (N. California), and EU (Ireland). Your manager instructed you to set up a latency-based routing to route incoming traffic for www.tutorialsdojo.com to all the EC2 instances across all AWS regions. Which of the following options can satisfy the given requirement? R: Use Route 53 to distribute the load to the multiple EC2 instances across all AWS Regions.

3. You are working for a startup which develops an AI-based traffic monitoring service. You need to register a new domain called www.tutorialsdojo-ai.com and set up other DNS entries for the other components of your system

in AWS. Which of the following is not supported by Amazon Route 53? **R: DNSSEC (Domain Name System Security Extensions)**

4. Your company is planning to use Route 53 as the DNS provider. There is a need to ensure that the company's domain name points to an existing CloudFront distribution. How could this be achieved? R:
Create an Alias record which points to the CloudFront distribution.

5. Which of the following are correct options for logging and monitoring AWS Route 53 service? Choose 3. R: 1) Amazon CloudWatch 2) AWS Route 53 dashboard 3) AWS CloudTrail

6. You are planning to launch a web based application in two different regions within US on AWS due to your organization compliance policies. You have setup 2 EC2 instances attached to an elastic load balancer in us-east-1. You have replicated the same setup in us-west-1. Now you have two load balancers which needs to listen traffic from internet. You would want to split the requests equally between both load balancers from a domain name hosted on your AWS Route 53. How should you configure your Route 53 record sets? R: Create two record sets, one each for us-east-1 and us-west-1 load balancers. Set weighted routing policy with weights as 1 and 1 respectively.

7. You have an Amazon Route 53 alias record that routes the traffic to an Application Load Balancer. Later on, the availability zones enabled for the load balancer are changed by a team member. When you check the load balancer using the dig command, you find that the IPs of the ELB have changed. What kind of change do you need to do for the alias record in Route 53? **R: Nothing as Route 53 automatically recognizes changes in the resource for the alias record.**

8. Você está trabalhando como arquiteto da AWS para uma empresa iniciante. A empresa possui servidores web implantados em todas as AZ na região AWS eu-central-1 (Frankfurt). Esses servidores da web fornecem notícias para usuários da Alemanha. O aplicativo é implantado em vários servidores EC2 com vários endereços IP estáticos e você precisa criar um conjunto de registros para o aplicativo. Como você configuraria o conjunto de registros no Route 53? R: Multivalue answer routing policy. When Route 53 is configured with Multi-value answer routing, it returns multiple values for web-servers. Route 53 responds to DNS queries with up to eight healthy records. Traffic is approximately load-balanced between these multiple web-servers.

AWS Transit Gateway

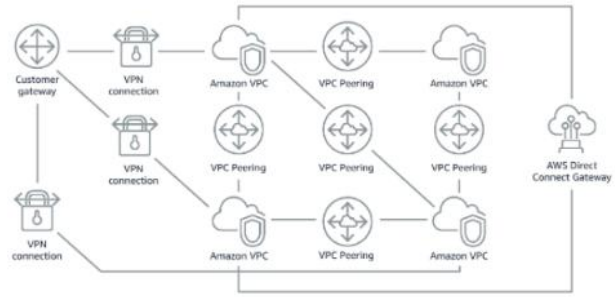
Conecte as Amazon VPCs, as contas da AWS e as redes locais a um único gateway

O AWS Transit Gateway conecta VPCs e suas redes locais por meio de um hub central. Isso simplifica a rede e elimina os complexos relacionamentos de emparelhamento. Ele atua como um roteador de nuvem e cada nova conexão só é feita uma vez.

À medida que você se expande globalmente, o emparelhamento entre regiões conecta os AWS Transit Gateways usando a rede global da AWS. Seus dados são automaticamente criptografados e nunca trafegam pela Internet pública.

E, devido à sua posição central, o AWS Transit Gateway Network Manager tem uma visão unificada de toda a rede, conectando até mesmo a dispositivos da Software-Defined Wide Area Network (SD-WAN – rede de área ampla definida por software).

Sem o AWS Transit Gateway



A complexidade cresce com a expansão. Você precisa manter as tabelas de roteamento em cada VPC e conectar a cada ponto local usando gateways de rede separados.

Com o AWS Transit Gateway

Sua rede é agilizada e escalável. O AWS Transit Gateway roteia todo o tráfego de e para cada VPC ou VPN, e você só tem um lugar para gerenciar e monitorar todas as operações.

Benefícios

Monitoramento de rede centralizado: O gerenciador de rede do AWS Transit Gateway inclui eventos e métricas para monitorar a qualidade da sua rede global, na AWS e on-premises. Os alertas de eventos especificam alterações no status da topologia, do roteamento e da conexão. As métricas de uso fornecem informações sobre conexões ativas/inativas, entrada/saída de bytes, entrada/saída de pacotes e pacotes descartados.

Visibilidade da rede global: Visualize e monitore sua rede global apenas no painel do gerenciador de rede do AWS Transit Gateway. Veja exibições em lista, lógica e mapa dos recursos e da conectividade da rede. O gerenciador de rede do Transit Gateway notifica sobre conexões não íntegras, alterações na disponibilidade e performance nas regiões da AWS e nos sites on-premises.

Integração da SD-WAN: O gerenciador de rede do AWS Transit Gateway se integra perfeitamente às soluções de SD-WAN da Cisco, Aruba, Silver Peak, Aviatrix e Versa Networks, tornando-a a interface unificada para gerenciar sua rede global na AWS e em localidades on-premises. Esses consoles de gerenciamento de SD-WAN são configurados para criar conexões AWS Site-to-Site VPN automaticamente entre suas instalações e a AWS.

Recursos

Roteamento: O AWS Transit Gateways suporta o roteamento dinâmico e estático da camada três entre Amazon Virtual Private Clouds (VPCs) e a VPN. As rotas determinam o próximo salto, dependendo do endereço IP de destino do pacote, e podem apontar para um Amazon VPC ou para uma conexão VPN.

Conectividade de Borda: Você pode criar conexões VPN entre o AWS Transit Gateway e os gateways locais usando a VPN. Você pode criar várias conexões VPN que anunciam os mesmos prefixos e habilitar o Equal Cost Multipath (ECMP) entre essas conexões. O ECMP consegue aumentar a largura de banda pelo tráfego de balanceamento de carga em vários caminhos.

Transit Gateway Connect: O AWS Transit Gateway Connect permite uma integração nativa dos dispositivos SD-WAN (Software-Defined Wide Area Network) na AWS. Os clientes agora podem estender perfeitamente sua borda SD-WAN na AWS usando protocolos padrão, como Generic Routing Encapsulation (GRE) e Border Gateway Protocol (BGP). Ele oferece aos clientes benefícios adicionais, como largura de banda aprimorada e suporte ao roteamento dinâmico com limites de rota aumentados, eliminando assim a necessidade de configurar várias VPNs IPsec entre os dispositivos SD-WAN e o Transit Gateway.

Interoperabilidade do Recurso Amazon VPC: O AWS Transit Gateway permite a resolução de nomes de host DNS públicos para endereços IP privados quando consultados em Amazon VPCs que também estão conectados ao AWS Transit Gateway. Uma instância em um Amazon VPC pode acessar um gateway NAT, o Network Load Balancer, o AWS PrivateLink e o Amazon Elastic File System em outros Amazon VPCs que também estão conectados ao AWS Transit Gateway.

Gerenciamento: Você pode usar a interface da linha de comandos (CLI), o Console de Gerenciamento da AWS ou o AWS CloudFormation para criar e gerenciar seu AWS Transit Gateway. O AWS Transit Gateway fornece métricas do Amazon CloudWatch, como o número de bytes enviados e recebidos entre os Amazon VCPs e as VPNs, a contagem de pacotes e a contagem de descartes. Além disso, você pode usar os logs de fluxo do Amazon VPC com o AWS Transit Gateway para coletar informações sobre o tráfego IP que passa pela associação do AWS Transit Gateway.

Amazon VPC (and associated features)

A Amazon Virtual Private Cloud (Amazon VPC) é um serviço que permite iniciar recursos da AWS em uma rede virtual isolada logicamente definida por você

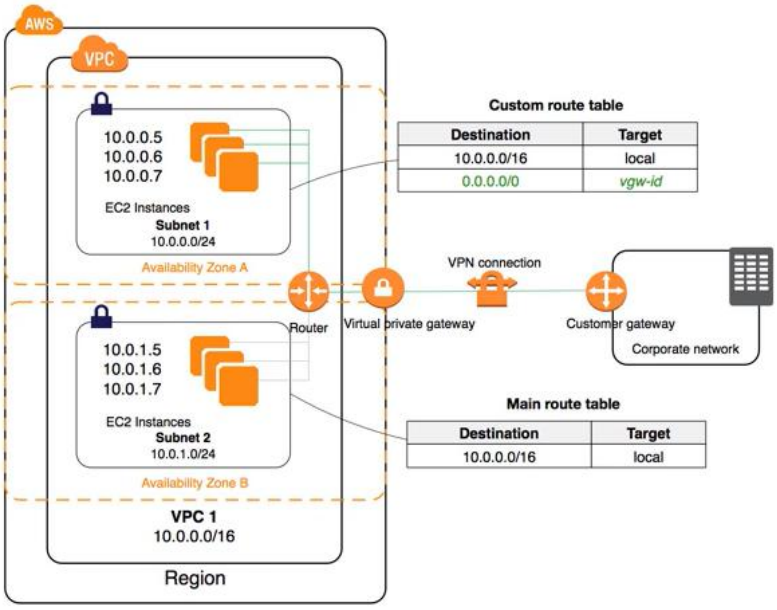
Você tem controle total sobre seu ambiente de redes virtuais, incluindo a seleção do seu próprio intervalo de endereços IP, a criação de sub-redes e a configuração de tabelas de rotas e gateways de rede

Você pode usar IPv4 e IPv6 para a maioria dos recursos em sua nuvem privada virtual, garantindo acesso fácil e seguro a recursos e aplicações

Como um dos serviços básicos da AWS, o Amazon VPC facilita a personalização da configuração de rede da VPC. Você pode criar uma sub-rede voltada ao público para seus servidores Web que têm acesso à Internet

Também é possível colocar seus sistemas back-end, como bancos de dados ou servidores de aplicações, em uma sub-rede privada, sem acesso à Internet

Com o Amazon VPC, você pode usar várias camadas de segurança, incluindo grupos de segurança e listas de controle de acesso à rede, para ajudar a controlar o acesso às instâncias do Amazon EC2 em cada sub-rede

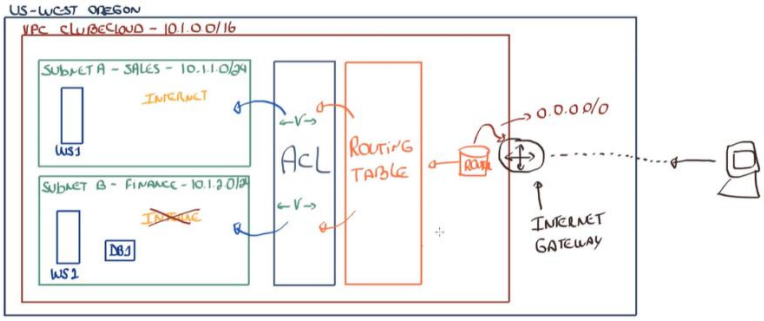


Trocando em Miúdos

É possível criar um **Internet Gateway** para uma determinada sub rede para que ela tenha acesso externo a internet

Podemos ter por exemplo uma sub rede A com instâncias EC2 e outra sub rede B com instâncias de database onde a sub rede B não tem acesso a internet tanto para enviar quanto para receber dados. As EC2 se comunicarão normalmente com a sub rede B, porém a sub rede B estará bem mais segura contra ataques e falhas de segurança por exemplo

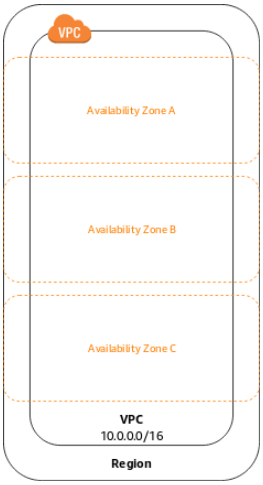
É possível também fazer com que a VPC pareça estar dentro da empresa através de um **Virtual Private Gateway**. Ou seja, podemos criar uma VPN entre a empresa e a VPC dentro da AWS para que se comuniquem



Noções básicas de VPC e sub-rede

Uma nuvem privada virtual (VPC) é uma rede virtual dedicada à sua conta AWS. Ele é logicamente isolado de outras redes virtuais na nuvem AWS. Você pode iniciar seus recursos da AWS, como instâncias do Amazon EC2, em seu VPC.

Ao criar um VPC, você deve especificar um intervalo de endereços IPv4 para o VPC na forma de um bloco Classless Inter-Domain Routing (CIDR); por exemplo, 10.0.0.0/16. Este é o bloco CIDR primário para seu VPC. O diagrama a seguir mostra um novo VPC com um bloco CIDR IPv4.



Default VPC vs Custom VPC

There is a limit of 5 VPCs per region

Quando criamos instâncias EC2, automaticamente elas são adicionadas a uma VPC default

Default VPC são mais amigáveis, pois proporciona algumas features default a determinados serviços. Temos como exemplo a criação de instâncias EC2 que recebem através da VPC default um IP privado, IP público e acesso à internet

A Default VPC por ser automático o usuário não tem tanto controle, porém na Custom VPC por ser manual, o usuário ganha mais controle

Custom VPC praticamente tudo tem que ser criado como:

- Subnet
- IP privado
- IP publico
- Roteamento

VPC Peering

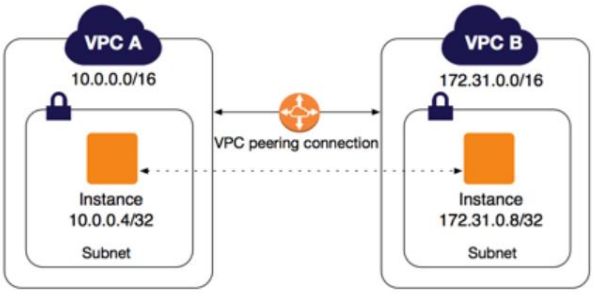
A VPC peering connection is a networking connection between two VPCs that enable you to route traffic between them privately. Instances in either VPC can communicate with each other as if they are within the same network. You can create a VPC peering connection between your own VPCs, with a VPC in another AWS account, or with a VPC in a different AWS Region.

Uma conexão de emparelhamento de VPC é uma conexão de rede entre duas VPCs que permite direcionar o tráfego entre elas usando endereços IPv4 ou IPv6 privados. Instâncias em qualquer VPC podem se comunicar umas com as outras como se estivessem na mesma rede.

É possível criar uma conexão de emparelhamento de VPC entre suas próprias VPCs ou com uma VPC de outra conta da AWS. As VPCs podem estar em regiões diferentes (também conhecidas como conexão de emparelhamento de VPC entre regiões).

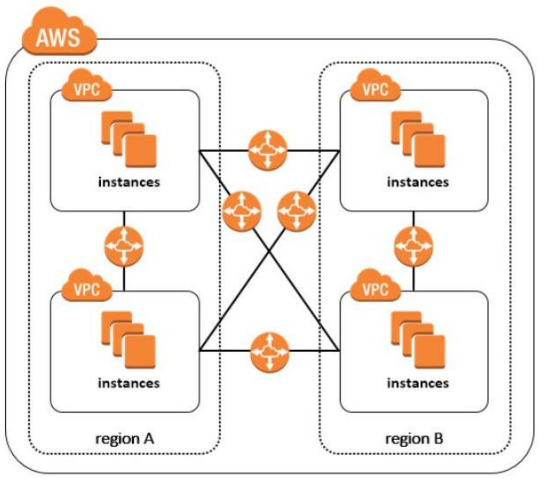
Importante: VPC Peering, no máximo 4 conexões simultâneas por VPC (uma VPC conectada em outras 4 VPCs)

Observação: Não há *Transit* entre as VPCs. Suponha que a VPC 1 esteja conectada a VPC 2 que por sua vez está conecta a VPC 3, a VPC 1 não conseguira “transitar” para a VPC 3, ou seja uma conexão por VPC



A AWS usa a infraestrutura existente de uma VPC para criar uma conexão de emparelhamento de VPC; não é nem um gateway nem uma conexão VPN, e não depende de uma parte separada do hardware físico. Não há um ponto único de falha de comunicação ou um gargalo de largura de banda

O tráfego sempre fica na estrutura global da AWS e nunca atravessa a Internet pública, o que reduz ameaças, como violações comuns e ataques de DDoS



Router Table

Uma tabela de rotas contém um conjunto de regras, chamadas rotas, que são usadas para determinar para onde o tráfego de rede de sua sub-rede ou gateway é direcionado.

Security Groups

Um grupo de segurança atua como um firewall virtual para sua instância para controlar o tráfego de entrada e saída. Ao iniciar uma instância em um VPC, você pode atribuir até cinco grupos de segurança à instância.

Os grupos de segurança atuam no nível da instância, não no nível da sub-rede. Portanto, cada instância em uma sub-rede em seu VPC pode ser atribuída a um conjunto diferente de grupos de segurança.

Se você iniciar uma instância usando a API Amazon EC2 ou uma ferramenta de linha de comando e não especificar um grupo de segurança, a instância será automaticamente atribuída ao grupo de segurança padrão para o VPC.

Se você iniciar uma instância usando o console do Amazon EC2, terá a opção de criar um novo grupo de segurança para a instância.

Para cada grupo de segurança, você adiciona regras que controlam o tráfego de entrada para as instâncias e um conjunto separado de regras que controlam o tráfego de saída.

ACL vs Security Groups

É possível adicionar uma ACL para filtrar o trafego para a VPC e para a Subnet. Por exemplo, é possível permitir que determinados IPs tenham acesso à internet e o restante não tenha acesso

Uma ACL por definição é Stateless, significa que é possível criar regras de Allow e regras de Deny

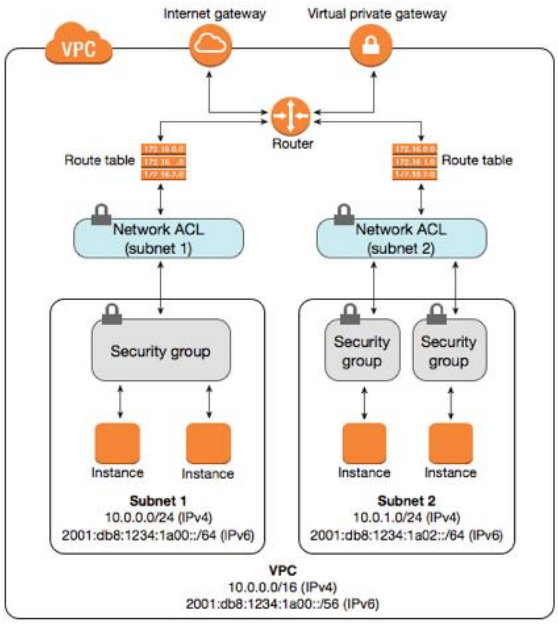
Security Groups por definição é Stateful, significa que só é possível permitir, o restante é bloqueado

Security Groups têm estado - se você enviar uma solicitação de sua instância, o tráfego de resposta para essa solicitação poderá fluir independentemente das regras do grupo de segurança de entrada. As respostas ao tráfego de entrada permitido podem fluir para fora, independentemente das regras de saída.

As ACLs de rede são sem estado; as respostas ao tráfego de entrada permitido estão sujeitas às regras para o tráfego de saída (e vice-versa).

By default, Amazon VPC security groups allow all outbound traffic unless you've specifically restricted outbound access. For a gateway endpoint, if your security group's outbound rules are restricted, you must add a rule that allows outbound traffic from your VPC to the service that's specified in your endpoint. To do this, you can use the service's prefix list ID as the destination in the outbound rule.

The default network ACL is configured to allow all traffic to flow in and out of the subnets with which it is associated. If your network ACL rules restrict traffic, you must specify the CIDR block (IP address range) for Amazon S3. So this option is correct.



Diferença entre Endereços Public e Private

Para os endereços IP privados, a AWS utiliza prefixo (Prefix Lenth) para representar uma máscara de sub-rede (exemplo: 10.1.1.2/24 representa a máscara 255.255.255.0 / 8 + 8 + 8 + 0 = 24

Lembrando que uma máscara de rede 255.255.255.0 os 3 primeiros octetos representam a rede e o ultimo octeto o host

Privados

- Classe A → 10.0.0.0 até 10.255.255.255
- Classe B → 172.16.0.0 até 172.31.255.255
- Classe C → 192.168.0.0 até 192.168.255.255

Públicos

Todos os IPs que não estão dentro das classes de IPs privados são públicos, exemplo: 11.1.1.1 (público)

Disponibilidade de IPs

Quando se cria uma Subnet na AWS, como por exemplo 10.1.1.0/24, haverá 251 IPs disponíveis (256 - 5 = 251)

- Isso porque o primeiro IP (10.1.1.0) será o da própria rede e não pode ser utilizado
- O último IP (10.1.1.255) não pode ser utilizado pois é reservado para broadcast
- O IP 10.1.1.1 será reservado para o roteador VPC
- O IP 10.1.1.2 será reservado para o servidor de DNS
- O IP 10.1.1.3 será reservado para uso futuro

The first four IP addresses and the last IP address in each subnet CIDR block are not available for you to use, and cannot be assigned to an instance. For example, in a subnet with CIDR block `10.0.0.0/24`, the following five IP addresses are reserved:

- `10.0.0.0`: Network address.
- `10.0.0.1`: Reserved by AWS for the VPC router.
- `10.0.0.2`: Reserved by AWS. The IP address of the DNS server is always the base of the VPC network range plus two; however, we also reserve the base of each subnet range plus two. For VPCs with multiple CIDR blocks, the IP address of the DNS server is located in the primary CIDR. For more information, see Amazon DNS Server.
- `10.0.0.3`: Reserved by AWS for future use.
- `10.0.0.255`: Network broadcast address. We do not support broadcast in a VPC, therefore we reserve this address.

DNS Hostname

By default, custom VPCs does not have DNS Hostnames enabled. So when you launch an EC2 instance in custom VPC, you do not have a public DNS name. You should go to VPC actions € Edit DNS Hostnames and enable it to have DNS hostnames for the resources within VPC.

DNS Support in Your VPC

Your VPC has attributes that determine whether your instance receives public DNS hostnames, and whether DNS resolution through the Amazon DNS server is supported.

Attribute	Description
`enableDnsHostnames`	Indicates whether the instances launched in the VPC get public DNS hostnames. If this attribute is `true`, instances in the VPC get public DNS hostnames, but only if the `enableDnsSupport` attribute is also set to `true`.
`enableDnsSupport`	Indicates whether the DNS resolution is supported for the VPC. If this attribute is `false`, the Amazon-provided DNS server in the VPC that resolves public DNS hostnames to IP addresses is not enabled. If this attribute is `true`, queries to the Amazon provided DNS server at the 169.254.169.253 IP address, or the reserved IP address at the base of the VPC IPv4 network range plus two will succeed. For more information, see Amazon DNS Server.

Como funciona o NAT

NAT é realizada em um computador que é um gateway entre duas redes entre uma empresa privada rede e da Internet, por exemplo. O computador em que a NAT é realizado é um **servidor proxy**. Ele pode realizar outros serviços de proxy, como filtragem de conteúdo ao mesmo tempo que faz o re-endereçamento de pacotes de dados

Sem o NAT a sua rede privada não conseguiria se comunicar com a internet, simplesmente por que a internet não iria reconhecer o IP da sua rede privada

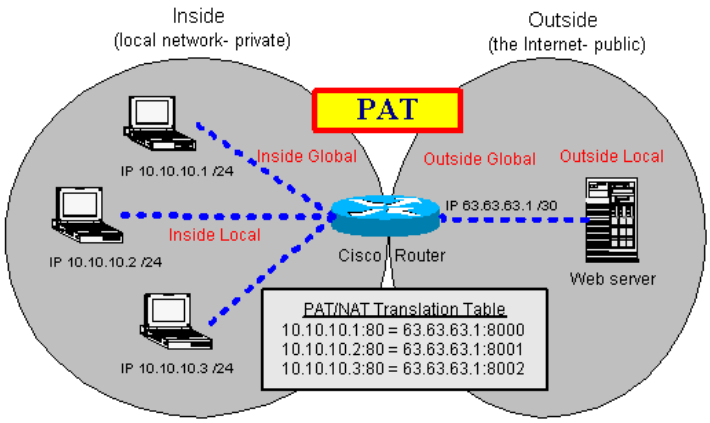
PAT (Port Address Translation) é uma característica de um dispositivo de rede que uma porta ou IP é ligada a um IP de uma LAN. Basicamente o que ele faz é associar o IP público e privado a uma porta (um número), por exemplo: 10.1.1.81:1 -> 200.1.1.1:1

Quando é NAT? Quando é 1 IP de internet para 1 IP de rede privada
Quando é PAT? PAT (Port Address Translation) é quando você tem 1 único IP valido para internet e tem vários clientes ou maquinas atrás desse único IP de NAT. Desta forma ele deixa de ser um NAT e passa a ser chamado de PAT

Mas porque o nome PAT?

Devido a ele traduzir portas... Para tornar possível vários IPs de rede privada usarem um único IP de internet ele traduz portas e isto é possível até ele usar todas as 65535 portas disponíveis.

Explicando melhor. Tente imaginar 2 maquinas com IPs de rede privada saindo para internet para um mesmo site. O PAT traduz qual maquina entrou em qual site e manda os pacotes corretamente para as maquinas que fizeram a requisição. Lembre-se que essas maquinas saíram com o mesmo IP para o mesmo site, porém são maquinas diferentes internamente



Não deixa de ser NAT, pois a tradução de IP continua. Só há PAT quando há forward explicito, ou seja, regras definidas para isto. Quando não há regras definidas a comunicação da tradução se dá de forma via socket entre os hosts que saem apenas com um IP válido atrás do nateamento. *"Socket TCP identificado pela 4-dupla: endereço IP origem; número da porta origem; endereço IP destino; número do porta destino"*

O que o NAT/PAT faz é pegar o IP privado da rede, por exemplo, 10.1.1.81 e traduzir para o IP público 200.1.1.1 (o destino é 8.8.8.8), e quando o pacote que foi enviado pelo destino 8.8.8.8 é retornado para o IP de origem (IP público), o NAT/PAT guardou na sua tabela de roteamento quem fez a solicitação para o IP 8.8.8.8 e retorna o pacote para o IP 10.1.1.81. Quando é PAT, o roteador consegue saber pela porta que está explicitada no pacote e consegue direcionar para o IP + Porta correspondente

Dentro da topologia AWS, o primeiro sendo chamado de NAT instance e o segundo sendo chamado de NAT Gateway. O NAT instance é um servidor EC2 e o segundo é uma aplicação

NAT Gateway

Para criar um gateway NAT, você deve especificar uma sub-rede pública na qual o gateway NAT residirá. Para obter mais informações sobre sub-redes públicas e privadas, consulte Roteamento de sub-rede

Também é necessário especificar um endereço IP elástico para associá-lo com o gateway NAT ao criá-lo. O endereço IP elástico não poderá ser alterado depois que for associado ao gateway NAT

Depois que criar um gateway NAT, será necessário atualizar a tabela de rotas associada a uma ou mais de suas sub-redes privadas para direcionar o tráfego vinculado à Internet para o gateway NAT. Isso permite que as instâncias nas sub-redes privadas se comuniquem com a internet

Todo gateway NAT é criado em uma Zona de disponibilidade específica e implementado com redundância nessa zona. Existe uma cota de gateways NAT que podem ser criados em uma zona de disponibilidade. Para obter mais informações, consulte Cotas da Amazon VPC

Se você tiver recursos em várias zonas de disponibilidade e eles compartilharem um gateway NAT, caso a zona de disponibilidade do gateway NAT fique inativa, os recursos em outras zonas de disponibilidade perderão o acesso à Internet

Para criar uma Zona de disponibilidade com arquitetura independente, crie um gateway NAT em cada Zona de disponibilidade e configure seu roteamento para garantir que os recursos usem o gateway NAT na mesma Zona de disponibilidade.

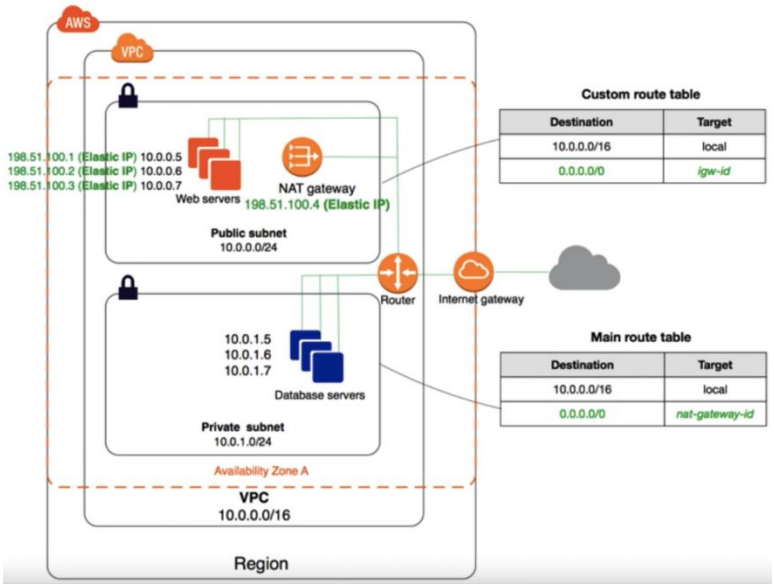
Caso não precise mais de um gateway NAT, você pode excluí-lo. A exclusão de um gateway NAT dissocia o respectivo endereço IP elástico, mas não libera o endereço de sua conta.

NAT Gateway cannot be created without an elastic IP address. During the creation of NAT Gateway, Elastic IP Allocation ID is a mandatory field without which we cannot proceed to create NAT Gateway.

When creating NAT Gateway, there is an option to select a subnet in which NAT Gateway will be created. This must be a public subnet that has a route to the internet through Internet Gateway.

Internet Gateway vs NAT Gateway: Quando uma Subnet utiliza o Internet Gateway ele ganha um IP público para sair para a internet diferentemente do NAT Gateway onde a Subnet não conhece o IP público, mas mesmo assim consegue sair para a internet

An IGW to a VPC allows instances with public IPs to access the internet. In contrast, NATs allow instances with no public IPs to access the internet.



Network ACL

Uma lista de controle de acesso à rede (ACL) é uma camada opcional de segurança para seu VPC que atua como um firewall para controlar o tráfego de entrada e saída de uma ou mais sub-redes. Você pode configurar ACLs de rede com regras semelhantes aos seus grupos de segurança para adicionar uma camada adicional de segurança ao seu VPC.

A seguir estão os itens básicos que você precisa saber sobre ACLs de rede:

- Seu VPC vem automaticamente com uma ACL de rede padrão modificável. Por padrão, ele permite todo o tráfego IPv4 de entrada e saída e, se aplicável, o tráfego IPv6.
- Você pode criar uma ACL de rede personalizada e associá-la a uma sub-rede. Por padrão, cada ACL de rede personalizada nega todo o tráfego de entrada e saída até que você adicione regras.

- Cada sub-rede em seu VPC deve ser associada a uma rede ACL. Se você não associar explicitamente uma sub-rede a uma ACL de rede, a sub-rede será automaticamente associada à ACL de rede padrão.
- Você pode associar uma ACL de rede a várias sub-redes. No entanto, uma sub-rede pode ser associada a apenas uma ACL de rede por vez. Quando você associa uma ACL de rede a uma sub-rede, a associação anterior é removida.
- Uma rede ACL contém uma lista numerada de regras. Avaliamos as regras em ordem, começando com a regra de menor número, para determinar se o tráfego é permitido dentro ou fora de qualquer sub-rede associada à rede ACL. O número mais alto que você pode usar para uma regra é 32766. Recomendamos que você comece criando regras em incrementos (por exemplo, incrementos de 10 ou 100) para que possa inserir novas regras onde precisar mais tarde.
- Uma rede ACL tem regras de entrada e saída separadas, e cada regra pode permitir ou negar o tráfego.
- As ACLs de rede não têm estado, o que significa que as respostas ao tráfego de entrada permitido estão sujeitas às regras para o tráfego de saída (e vice-versa).
- Para que o NAT gateway tenha acesso externo, tanto o Inbound rules quanto Outbound rules devem permitir o trafego para o range de portas TCP entre 1024 a 65535

Flow Logs

VPC Flow Logs is a feature that enables you to capture information about the IP traffic going to and from network interfaces in your VPC. Flow log data can be published to Amazon CloudWatch Logs and Amazon S3. After you've created a flow log, you can retrieve and view its data in the chosen destination.

You can create a flow log for a:

- VPC
- subnet
- network interface

Os registros de fluxo de VPC são um recurso que permite capturar informações sobre o tráfego IP que vai de e para as interfaces de rede em seu VPC. Os dados do log de fluxo podem ser publicados no Amazon CloudWatch Logs ou Amazon S3. Depois de criar um registro de fluxo, você pode recuperar e visualizar seus dados no destino escolhido.

Os registros de fluxo podem ajudá-lo com uma série de tarefas, como:

- Diagnosticar regras de grupo de segurança excessivamente restritivas
- Monitorar o tráfego que está alcançando sua instância
- Determinar a direção do tráfego de e para as interfaces de rede

Os dados do log de fluxo são coletados fora do caminho de seu tráfego de rede e, portanto, não afetam o rendimento ou a latência da rede. Você pode criar ou excluir logs de fluxo sem nenhum risco de impacto no desempenho da rede.

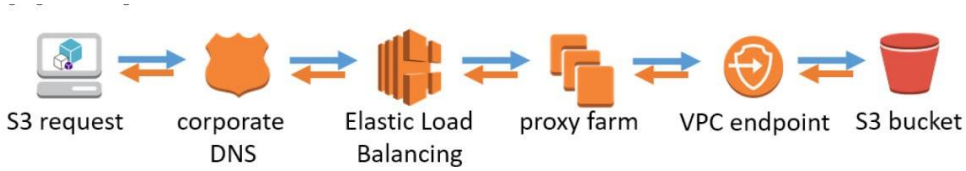
VPC End Point

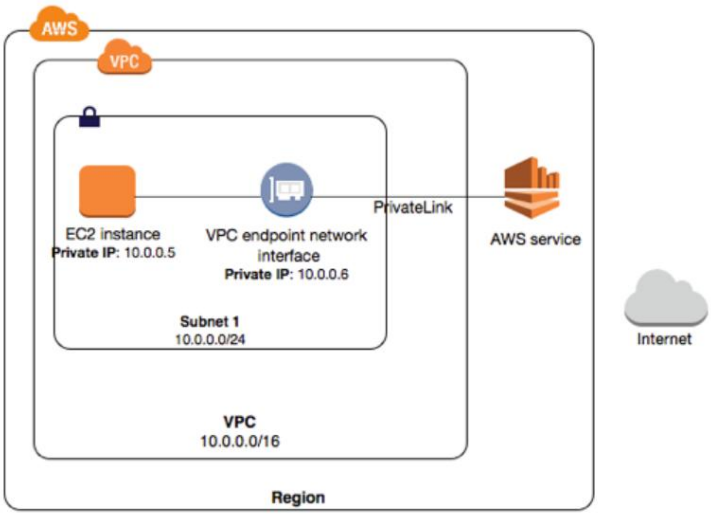
Quando se está dentro de uma VPC, não há como ter acesso aos serviços como S3, SNS, SQS, dentre outros. Para acessar os recursos da AWS é necessário acessar por fora, ou seja, sair da AWS pela internet e voltar para a infra da AWS para poder se conectar aos serviços.

Para resolver este problema é necessário utilizar o recurso de End Point. Ele cria uma espécie de túnel entre a VPC e os serviços AWS, sem precisar sair pela internet, acontecendo tudo dentro da AWS.

You can have multiple endpoint routes to different services in a route table, and you can have multiple endpoint routes to the same service in different route tables. Still, you cannot have multiple endpoints to the same service in a single route table. For example, if you have two endpoints to Amazon S3 in your VPC, you cannot use the same route table for both endpoints.

Endpoint connections cannot be extended out of a VPC. VPC peering connection, AWS Direct Connect connection, or ClassicLink connection in your VPC cannot use the endpoint to communicate with resources in the endpoint service. To support such use cases, we can set up an S3 proxy server on AWS EC2 instance, as shown below.





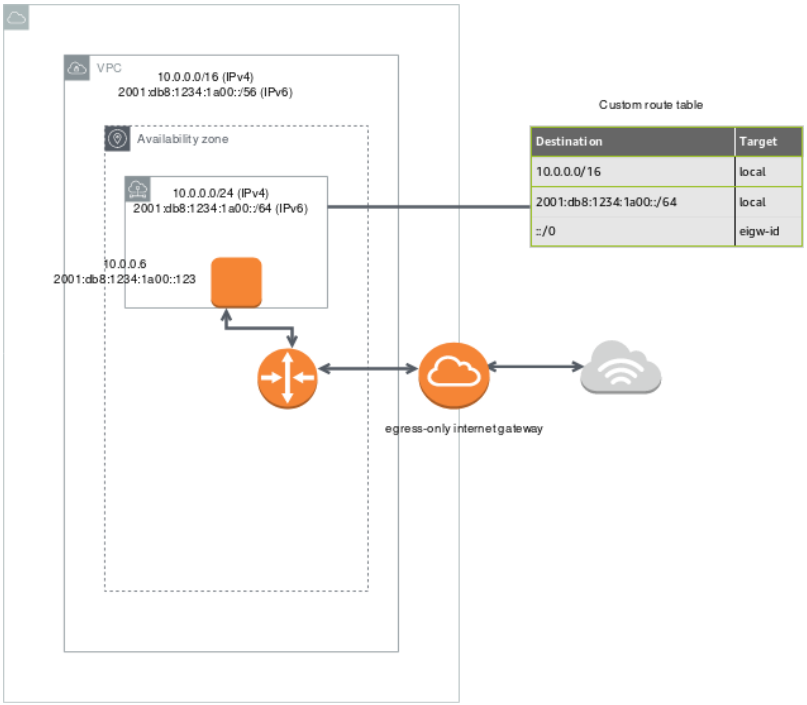
Egress-only internet gateways

An egress-only internet gateway is a horizontally scaled, redundant, and highly available VPC component that allows outbound communication over IPv6 from instances in your VPC to the internet, and prevents the internet from initiating an IPv6 connection with your instances.

An egress-only internet gateway is for use with IPv6 traffic only. To enable outbound-only internet communication over IPv4, use a NAT gateway instead.

IPv6 addresses are globally unique, and are therefore public by default. If you want your instance to be able to access the internet, but you want to prevent resources on the internet from initiating communication with your instance, you can use an egress-only internet gateway. To do this, create an egress-only internet gateway in your VPC, and then add a route to your route table that points all IPv6 traffic (::/0) or a specific range of IPv6 address to the egress-only internet gateway. IPv6 traffic in the subnet that's associated with the route table is routed to the egress-only internet gateway.

- You cannot associate a security group with an egress-only internet gateway. You can use security groups for your instances in the private subnet to control the traffic to and from those instances.
- You can use a network ACL to control the traffic to and from the subnet for which the egress-only internet gateway routes traffic.

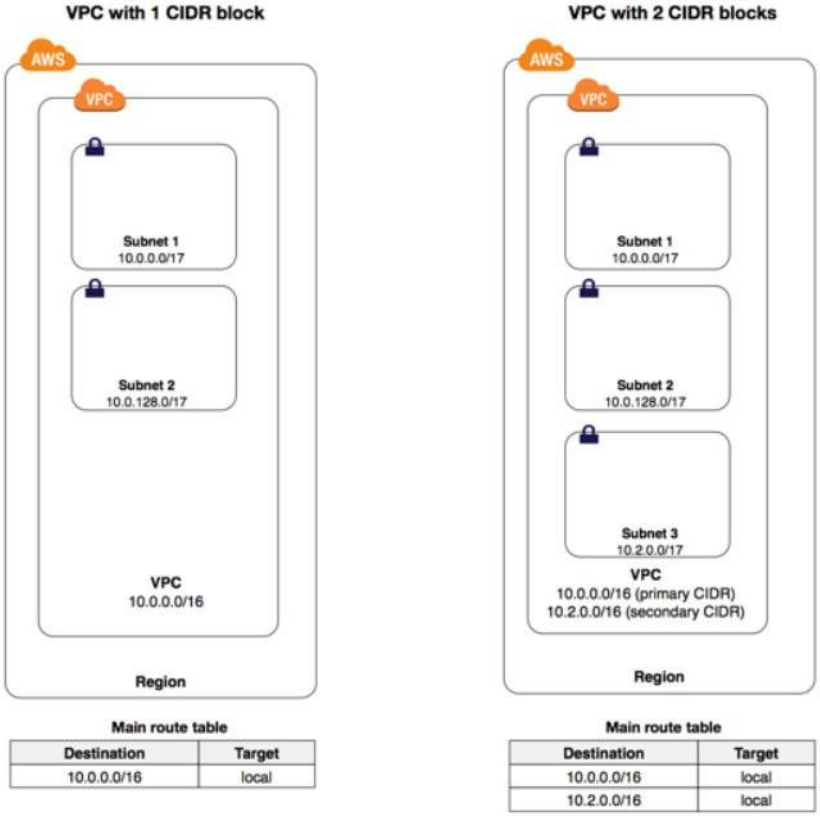


CIDR Block

O CIDR é uma sigla para Classes Inter-Domain Routing, e ele é considerado um método para repartir os endereços IP e para rotear.

You can associate secondary IPv4 CIDR blocks with your VPC. When you associate a CIDR block with your VPC, a route is automatically added to your VPC route tables to enable routing within the VPC (the destination is the CIDR block and the target is local).

In the following example, the VPC on the left has a single CIDR block (10.0.0.0/16) and two subnets. The VPC on the right represents the architecture of the same VPC after you've added a second CIDR block (10.2.0.0/16) and created a new subnet from the range of the second CIDR.



IP addresses to CIDR networks

Em seu AWS VPC, você precisa adicionar uma nova sub-rede que permitirá hospedar um total de 20 instâncias EC2. Qual bloco CIDR IPv4 você usaria para conseguir o mesmo? R: 151.0.0.0/27

AWS reserves 5 IP addresses.

A fórmula para calcular o número de endereços IP atribuíveis a redes CIDR é semelhante à rede classful. Subtraia o número de bits de rede de 32. Eleve 2 a essa potência e subtraia 2 para os endereços de rede e broadcast. Por exemplo, uma rede /24 tem 2 ^ (32-24) - 2 endereços disponíveis para atribuição de host.

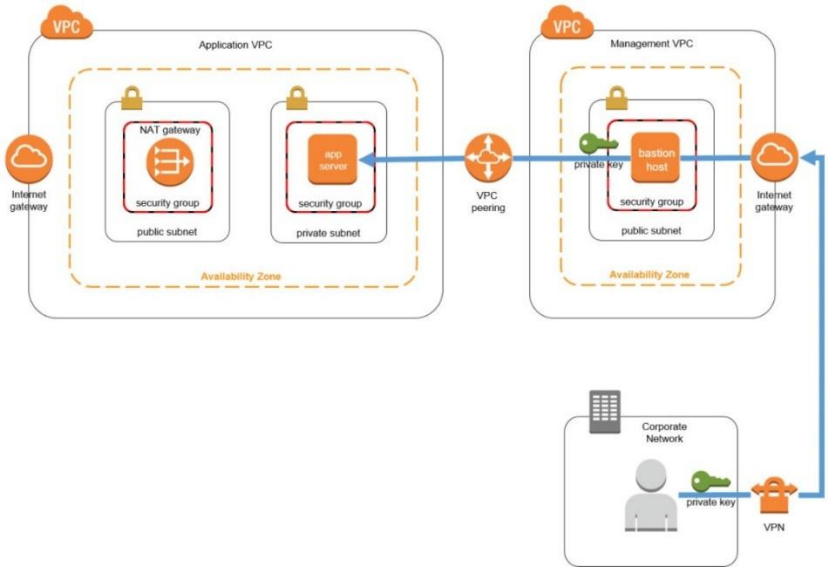
A. Prefix Length is '27'
Therefore 32-27 = 5 and 2 ^ 5 (i.e 2 * 2 * 2 * 2 * 2) – 5 = 27

B. Prefix Length is '28'
Therefore 32-28 = 4 and 2 ^ 4 (i.e 2 * 2 * 2 * 2) - 5= 11

C. Prefix Length is '29'
Therefore 32-29 = 3 and 2 ^ 3 (i.e 2 * 2 * 2) - 5 = 3

D. Prefix Length is '30'
Therefore 32-30 = 2 and 2 ^ 2 (i.e 2 * 2) - 5 = -1

Bastion Host



Custom Route Table

A custom route table can be made as to the main route table so that all implicit associations of subnets will now point to the newly set main route table. All the future implicit associations of newly created subnets will point to the newly set main route table.

Replacing the Main Route Table

You can change which route table is the main route table in your VPC.

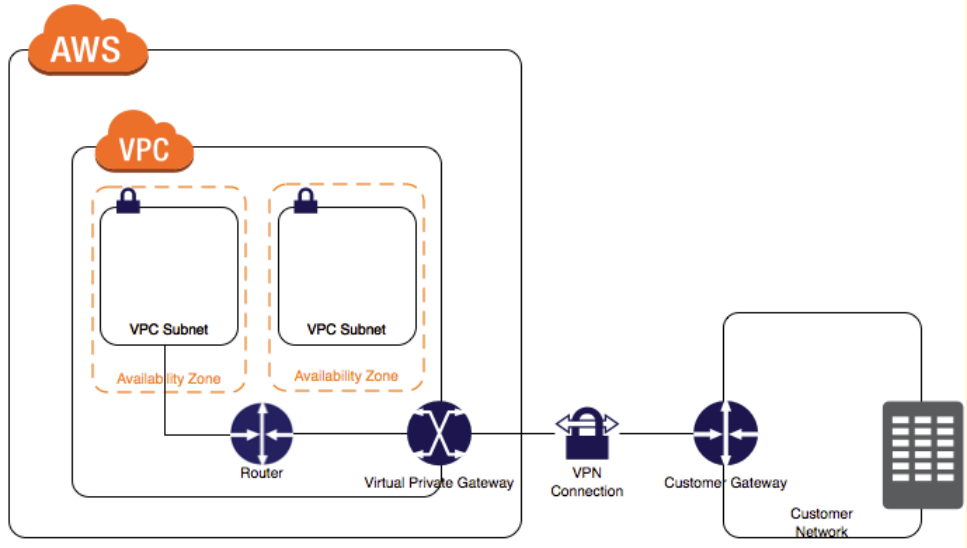
To replace the main route table

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose **Route Tables**.
3. Select the route table that should be the new main route table, and then choose **Set as Main Table**.
4. In the confirmation dialog box, choose **Yes, Set**.

The following procedure describes how to remove an explicit association between a subnet and the main route table. The result is an implicit association between the subnet and the main route table. The process is the same as disassociating any subnet from any route table.

AWS VPN connections

When defining a VPN connection between the on-premises network and the VPC, you need to have a customer gateway defined. Since this is accessed over the internet, it needs to have a static internet-routable IP Address.



All other options are invalid since this is only required for an AWS Direct Connect connection.

Você foi instruído a estabelecer uma conexão VPN site a site bem-sucedida de sua rede local para a VPC (nuvem privada virtual). Como arquiteto, qual dos seguintes pré-requisitos você deve garantir para estabelecer a conexão VPN site a site? Escolha 2 respostas das opções fornecidas abaixo.

R1: A public IP address on the customer gateway for the on-premises network

R2: A virtual private gateway attached to the VPC

Overlapping CIDR Blocks And Transitive Peering

Não é possível criar uma conexão de peering de VPC entre VPCs com blocos CIDR IPv4 correspondentes ou sobrepostos.

Overlapping CIDR Blocks

You cannot create a VPC peering connection between VPCs with matching or overlapping IPv4 CIDR blocks.

Transitive peering is unsupported for VPC Peering.

Links Úteis

https://docs.aws.amazon.com/vpc/latest/userguide/egress-only-internet-gateway.html

https://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-nat-comparison.html
https://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-nat-gateway.html#nat- gateway-basics
https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Internet_Gateway.html

➢ Pontos de Atenção

1. Uma grande seguradora tem uma conta AWS que contém três VPCs (DEV, UAT e PROD) na mesma região. O UAT é peering para PROD e DEV usando uma conexão de peering VPC. Todos os VPCs têm blocos CIDR não sobrepostos. A empresa deseja enviar versões de código menores do Dev para o Prod para acelerar o tempo de lançamento no mercado. Qual das opções a seguir ajuda a empresa a fazer isso? R: Crie uma nova conexão de peering VPC entre PROD e DEV com as rotas apropriadas.

2. Recentemente, você foi promovido a uma função de líder técnico em sua equipe de DevOps. Sua empresa tem um VPC existente que não é utilizado nos últimos meses. O gerente de negócios instruiu você a integrar seu data center local e seu VPC. Você explicou a lista de tarefas que executará e mencionou sobre uma conexão de rede privada virtual (VPN). O gerente de negócios não entende de tecnologia, mas está interessado em saber o que é uma VPN e seus benefícios. Qual é uma das principais vantagens de ter uma VPN na AWS? R: You can connect your AWS cloud resources to on-premises data centers using VPN connections.

3. Uma grande empresa financeira recentemente adotou uma arquitetura de nuvem híbrida para integrar seu data center local e sua nuvem AWS. Seu gerente o instruiu a criar uma nova topologia de rede VPC que deve oferecer suporte a todos os aplicativos da Web voltados para a Internet, bem como aos aplicativos voltados internamente que são acessados por funcionários apenas por VPN. Para garantir alta disponibilidade e tolerância a falhas, os aplicativos financeiros voltados para a Internet e internos devem ser capazes de aproveitar pelo menos duas AZs para alta disponibilidade. Qual é o número mínimo de sub-redes que você deve criar em seu VPC para acomodar esses requisitos? R: 4 subnets

4. Você está implementando uma arquitetura híbrida para sua empresa, em que conecta a Amazon Virtual Private Cloud (VPC) à rede local. Qual das opções a seguir pode ser usada para criar uma conexão privada entre o VPC e a rede local da sua empresa? R: Direct Connect

5. A VPC has a non-default public subnet which has four On-Demand EC2 instances that can be accessed over the Internet. Using the AWS CLI, you launched a fifth instance that uses the same subnet, Amazon Machine Image (AMI), and security group which are being used by the other instances. Upon testing, you are not able to access the new instance. Which of the following is the most suitable solution to solve this problem? R: Associate an Elastic IP address to the fifth EC2 instance.

6. You are working as a Senior Solutions Architect for a data analytics company which has a VPC for their human resource department, and another VPC for their finance department. You need to configure your architecture to allow the finance department to access all resources that are in the human resource department and vice versa.

Which type of networking connection in AWS should you set up to satisfy the above requirement? R: VPC Peering

7. You have designed and built a new AWS architecture. After deploying your application to an On-demand EC2 instance, you found that there is an issue in your application when connecting to port 443. After troubleshooting the issue, you added port 443 to the security group of the instance. How long will it take before the changes are applied to all of the resources in your VPC? R: Immediately.

8. You are working as a Solutions Architect in a well-funded financial startup. The CTO instructed you to launch a cryptocurrency mining server on a Reserved EC2 instance in us-east-1 region's private subnet which is using IPv6. Due to the financial data that the server contains, the system should be secured to avoid any unauthorized access and to meet the regulatory compliance requirements. In this scenario, which VPC feature allows the EC2 instance to communicate to the Internet but prevents inbound traffic? R: Egress-only Internet gateway

9. You are tasked to host a web application in a new VPC with private and public subnets. In order to do this, you will need to deploy a new MySQL database server and a fleet of EC2 instances to host the application. In which subnet should you launch the new database server into? R: The private subnet

10. You have a web application running on EC2 instances which processes sensitive financial information. All of the data are stored on an Amazon S3 bucket. The financial information is accessed by users over the Internet. The security team of the company is concerned that the Internet connectivity to Amazon S3 is a security risk. In this scenario, what will you do to resolve this security concern? R: Change the web architecture to access the financial data through a Gateway VPC Endpoint.

11. Your company has setup EC2 Instances in a VPC for their application. The IT Security department has advised that all traffic be monitored to the EC2 Instances. Which of the following features can be used to capture information for outgoing and incoming IP traffic from network interfaces in a VPC. R:
AWS VPC Flow Logs

12. You are working as an architect in your organization. You have peered VPC A as requester and VPC B as accepted, and both VPCs can communicate with each other. Now you want resources in the private subnets of both the VPCs to reach out to the internet. But anyone on the internet should not be able to reach resources within both the VPCs. Which of the below will achieve the desired outcome? R: Create NAT Gateways in both VPCs and configure routes for each VPC to use its own NAT Gateway

13. Your organization already had a VPC(10.10.0.0/16) setup with one public(10.10.1.0/24) and two private subnets – private subnet 1 (10.10.2.0/24) and private subnet 2 (10.10.3.0/24). The public subnet has the main route table, and two private subnets have two different route tables respectively. AWS sysops team reports a problem starting the EC2 instance in private subnet 1 cannot communicate to the RDS MySQL database on private subnet 2. What are the possible reasons? (choose 2 options) R: 1) RDS security group inbound rule is incorrectly configured with 10.10.1.0/24 instead of 10.10.2.0/24; 2) 10.10.3.0/24 subnet's NACL is modified to deny inbound on port 3306 from subnet 10.10.2.0/24

14. You have a bastion host EC2 instance on AWS VPC public subnet. You would want to SSH to Bastion host EC2 instance. What would be the secure and minimal configuration you need for SSH requests to work? Assume route table is already set up with Internet Gateway. R: Allow SSH protocol (port 22) on Security Group Inbound. Allow Network ACL inbound and Network ACL outbound for your IP address.

15. Your organization had asked to be cost-efficient in designing AWS solutions. You have created three VPCs (VPC A, VPC B, VPC C), peered VPC A to VPC B and VPC B to VPC C. You have created a NAT gateway in VPC B and would like to use the same NAT Gateway for resources within VPC A and VPC C. However, the resources within VPC A and VPC C cannot communicate to the internet through NAT Gateway, but resources in VPC B can communicate. What could be the reason? R: Using another VPC's NAT Gateway is not supported in AWS

16. You have set up a peering connection between two VPCs. You have launched EC2 instances in both VPCs and trying to communicate with each other through peering connections. However, you found the request is getting

timed out. From the following options, what could be the reason for the time out? (Select TWO) R: 1) Network ACLs have been configured not to allow traffic from peered VPC; 2) Route tables of both VPCs only contains specific IP range for peering connection and either of the EC2 instances does not belong to the configured IP ranges.

17. You have created a new VPC and a private subnet. You will also be setting up a VPN connection with your organization to communicate with resources within the VPC. Your organization would need DNS names for some of the on-premise applications to communicate with VPC resources. You have launched a new EC2 instance with Auto-assign public IP as enable. When the instance is ready to use, you found that the Public DNS name is missing. What should be done to enable it? R: Enable DNS Hostnames for VPC

18. You are taking over the AWS platform in your organization. You were asked to build a new application that would require a fleet of 20 EC2 instances inside a private VPC that should communicate with each other and no traffic going into the EC2 instances from the internet but should receive requests from all other EC2 instances inside the VPC. When you looked at the existing VPC, it was created with 10.10.0.0/24 CIDR range containing only 256 IP addresses. You noticed that 8 subnets were consuming all 256 IP addresses with /27 CIDR ranges. How can you change the CIDR range of the VPC? R: Add secondary CIDR range for the VPC

19. Your organization has a VPC set up with a custom route table having 40 routes for different use cases such as "VPC peering", "VPN connections", "NAT gateways" etc with different IP ranges. The Main route table had a local route to the internet gateway to act for the public subnet. Your VPC IP range is 10.10.0.0/16, and many teams are working on this. VPC needs to create different subnets for their respective applications that need a custom route table associated with it. However, many a time, these teams forget to associate the custom route table to the subnets explicitly. This leads to a lot of troubleshooting hours when the connections to the new subnets from the VPN do not work as expected. As an architect, how would you resolve this issue? R: Make the custom route table as the main route table. Any new subnets created will get associated with it implicitly.

20. You have set up two VPCs: VPC A has the address of "10.10.0.0/16". It also has a subnet with address space "10.10.1.0/24". VPC B has the address of "10.11.0.0/16". It also has a subnet with address space "10.11.1.0/28". You also have set up VPC peering connection between the two VPCs. What should be the respective route table entries in VPC A and VPC B? R: VPC B route table contains route with Destination as 10.10.1.0/24 and VPC A route table contains route with Destination as 10.11.1.0/28. *To send private IPv4 traffic from your instance to an instance in a peer VPC, you must add a route to the route table that's associated with your subnet in which your instance resides. The route points to the CIDR block (or a portion of the CIDR block) of the peer VPC in the VPC peering connection. The owner of the other VPC in the peering connection must also add a route to their subnet's route table to direct traffic back to your VPC.*

21. You are working as Cloud Solutions Engineer in an IT firm, and the firm has set up multiple VPN connections. They want to provide secure communication between multiple sites using the AWS VPN Cloud Hub. Which statement is the most accurate in describing what you must do to set this up correctly? How do you connect multiple sites to a VPC? R: Create a virtual private gateway with multiple customer gateways, each with unique Border Gateway Protocol (BGP) Autonomous System Numbers (ASNs).

22. You are working in a College as a Cloud Technical Advisor, and your college was maintaining all its data locally where they felt security and redundancy issues. So, you suggested deploying the application in AWS and use a NoSQL database for their database. While deploying the servers in AWS, the team needs your suggestion for creating new Security Groups. Can you select which of the following Option given by the team is true? (Select 2) R: 1) The default rules in a security group disallows all incoming traffic. 2) By default, outbound traffic is allowed

23. A 50 year old Computer Solutions company has a very big application that needs to be deployed to the AWS cloud from its existing server. The application is media access control (MAC) address dependent as per the application licensing terms. This application will be deployed in an on-demand EC2 instance with instance type r4.2xlarge. In this scenario, how can you ensure that the MAC address of the EC2 instance will not change even if the instance is restarted or rebooted? R: Use a VPC with an **elastic network interface** that has a fixed MAC Address

24. One of your colleagues, who is new to the company where you work as a cloud Architect, has some issues with IP Addresses. He has created an Amazon VPC with an IPV4 CIDR block 10.0.0.0/24, but now there is a requirement of hosting a few more resources to that VPC. As per his knowledge, he is thinking of creating a new VPC with a greater range. Could you suggest to him a better way that should be reliable? R:
You can expand existing VPC by adding Secondary CIDR to your current VPC

25. Atualmente, sua empresa possui um conjunto de servidores da Web em uma sub-rede pública e servidores de banco de dados na sub-rede privada. Você precisa garantir que os administradores de seu ambiente local possam acessar os servidores de banco de dados. Qual das opções a seguir é uma maneira segura de acessar os servidores de banco de dados? R: Create a bastion host in the public subnet. Ask the IT administrators to log into the database servers via the bastion host. *Um host bastião é um servidor cujo objetivo é fornecer acesso a uma rede privada a partir de uma rede externa, como a Internet. Por causa de sua exposição a ataques potenciais, um host bastião deve minimizar as chances de penetração. Por exemplo, você pode usar um host bastião para reduzir o risco de permitir conexões SSH de uma rede externa para as instâncias do Linux iniciadas em uma sub-rede privada de sua Amazon Virtual Private Cloud (VPC).*

26. Uma empresa tem um conjunto de VPCs definidos na AWS. Eles precisam se conectar à rede local. Eles precisam garantir que todos os dados sejam criptografados em trânsito. Qual das opções a seguir você usaria para conectar os VPCs às redes locais? **R: VPN connections**

27. Você criou um novo VPC e uma sub-rede privada. Você também configurará uma conexão VPN com sua organização para se comunicar com recursos dentro do VPC. Sua organização precisaria de nomes DNS para alguns dos aplicativos locais para se comunicar com recursos VPC. Você iniciou uma nova instância EC2 com a atribuição automática de IP público como habilitado. Quando a instância está pronta para uso, você descobre que o nome DNS público está faltando. O que deve ser feito para habilitá-lo? R: Enable DNS Hostnames for VPC

28. Sua organização configurou um VPC com intervalo CIDR 10.10.0.0/16. Há um total de 100 sub-redes no VPC e estão sendo usadas ativamente por várias equipes de aplicativos. Uma equipe de aplicativo que está usando 50 instâncias EC2 na sub-rede 10.10.55.0/24 reclama que há falhas de conexão de rede de saída intermitente para cerca de 30 instâncias EC2 aleatórias em um determinado dia. Como você resolveria o problema com configuração mínima e registros mínimos gravados? R: Create a flow log for subnet 10.10.55.0/24.

29. Você está trabalhando como engenheiro de soluções em nuvem em uma empresa de TI, e a empresa configurou várias conexões VPN. Eles desejam fornecer comunicação segura entre vários sites usando o AWS VPN Cloud Hub. Qual afirmação é a mais precisa para descrever o que você deve fazer para configurar isso corretamente? Como você conecta vários sites a um VPC? **R: Create a virtual private gateway with multiple customer gateways, each with unique Border Gateway Protocol (BGP) Autonomous System Numbers (ASNs).** CORRECT because to use AWS VPN Cloud Hub, one must create a virtual private gateway with multiple customer gateways, each with a unique Border Gateway Protocol (BGP) Autonomous System Number (ASN).